The Annals of Statistics
2010, Vol. 38, No. 4, 2388–2421
DOI: 10.1214/09-AOS789

SIMULTANEOUS NONPARAMETRIC INFERENCE OF TIME SERIES[1]

By Weidong Liu and Wei Biao Wu

University of Pennsylvania and University of Chicago

We consider kernel estimation of marginal densities and regression functions of stationary processes. It is shown that for a wide class of time series, with proper centering and scaling, the maximum deviations of kernel density and regression estimates are asymptotically Gumbel. Our results substantially generalize earlier ones which were obtained under independence or beta mixing assumptions. The asymptotic results can be applied to assess patterns of marginal densities or regression functions via the construction of simultaneous confidence bands for which one can perform goodness-of-fit tests. As an application, we construct simultaneous confidence bands for drift and volatility functions in a dynamic short-term rate model for the U.S. Treasury yield curve rates data.

1. Introduction. Consider the nonparametric time series regression model

$$Y_i = \mu(X_i)\,dt + \sigma(X_i)\eta_i, \tag{1.1}$$

where $\mu(\cdot)$ [resp., $\sigma^2(\cdot)$] is an unknown regression (resp., conditional variance) function to be estimated, (X_i, Y_i) is a stationary process and η_i are unobserved independent and identically distributed (i.i.d.) errors with $\mathsf{E}\eta_i = 0$ and $\mathsf{E}\eta_i^2 = 1$. Let the regressor X_i be a stationarity causal process

$$X_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{1.2}$$

where ε_i are i.i.d. and the function G is such that X_i exists. Assume that η_i is independent of $(\ldots, \varepsilon_{i-1}, \varepsilon_i)$. Hence, η_i and $(\mu(X_i), \sigma(X_i))$ are independent. As a special case of (1.1), a particularly interesting example is the nonlinear autoregressive model

$$Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\eta_i, \tag{1.3}$$

Received August 2009; revised December 2009.
[1]Supported in part by NSF Grant DMS-04-78704.
AMS 2000 subject classifications. Primary 62H15; secondary 62G10.
Key words and phrases. Gumbel distribution, kernel density estimation, linear process, maximum deviation, nonlinear time series, nonparametric regression, simultaneous confidence band, stationary process, treasury bill data.

where $X_i = Y_{i-1}$ and $\varepsilon_i = \eta_{i-1}$. Many nonlinear time series models are of form (1.3) with different choices of $\mu(\cdot)$ and $\sigma(\cdot)$. If the form of $\mu(\cdot)$ is not known, we can use the Nadaraya–Watson estimator

$$\mu_n(x) = \frac{1}{nbf_n(x)} \sum_{k=1}^{n} K\left(\frac{X_k - x}{b}\right) Y_k, \tag{1.4}$$

where K is a kernel function with $K(\cdot) \geq 0$ and $\int_{\mathsf{R}} K(u)\,du = 1$, the bandwidths $b = b_n \to 0$ and $nb_n \to \infty$, and

$$f_n(x) = \frac{1}{nb} \sum_{k=1}^{n} K\left(\frac{X_k - x}{b}\right)$$

is the kernel density estimate of f, the marginal density of X_i. Asymptotic properties of nonparametric estimates for time series have been widely discussed under various strong mixing conditions; see Robinson (1983), Györfi et al. (1989), Tjøstheim (1994), Bosq (1996), Doukhan and Louhichi (1999) and Fan and Yao (2003), among others.

Under appropriate dependence conditions [see, e.g., Robinson (1983), Wu and Mielniczuk (2002), Fan and Yao (2003) and Wu (2005)], we have the central limit theorem

$$\sqrt{nb}[f_n(x) - \mathsf{E}f_n(x)] \Rightarrow N(0, \lambda_K f(x)) \qquad \text{where } \lambda_K = \int_{\mathsf{R}} K^2(u)\,du.$$

The above result can be used to construct point-wise confidence intervals of $f(x)$ at a fixed x. To assess shapes of density functions so that one can perform goodness-of-fit tests, however, one needs to construct *uniform* or *simultaneous confidence bands* (SCB). To this end, we need to deal with the maximum absolute deviation over some interval $[l, u]$:

$$\Delta_n := \sup_{l \leq x \leq u} \frac{\sqrt{nb}}{\sqrt{\lambda_K f(x)}} |f_n(x) - \mathsf{E}f_n(x)|. \tag{1.5}$$

In an influential paper, Bickel and Rosenblatt (1973) obtained an asymptotic distributional theory for Δ_n under the assumption that X_i are i.i.d. It is a very challenging problem to generalize their result to stationary processes where dependence is the rule rather than the exception. In their paper Bickel and Rosenblatt applied the very deep embedding theorem of approximating empirical processes of independent random variables by Brownian bridges with a reasonably sharp rate [Brillinger (1969), Komlós, Major and Tusnády (1975, 1976)]. For stationary processes, however, such an approximation with similar rates can be extremely difficult to obtain. Doukhan and Portal (1987) obtained a weak invariance principle for empirical distribution functions. In 1998, Neumann (1998) made a breakthrough and proved

a very useful result for β-mixing processes whose mixing rates decay exponentially quickly. Such processes are very weakly dependent. For mildly weakly dependent processes, the asymptotic problem of Δ_n remains open. Fan and Yao [(2003), page 208] conjectured that similar results hold for stationary processes under certain mixing conditions. Here we shall solve this open problem and establish an asymptotic theory for both short- and long-range dependent processes. It is shown that, for a wide class of short-range dependent processes, we can have a similar asymptotic distributional theory as Bickel and Rosenblatt (1973). However, for long-range dependent processes, the asymptotic behavior can be sharply different. One observes the dichotomy phenomenon: the asymptotic properties depend on the interplay between the strength of dependence and the size of bandwidths. For small bandwidths, the limiting distribution is the same as the one under independence. If the bandwidths are large, then the limiting distribution is half-normal [cf. (2.9)].

A closely related problem is to study the asymptotic uniform distributional theory for the Nadaraya–Watson estimator $\mu_n(x)$. Namely, one needs to find the asymptotic distribution for $\sup_{x\in T}|\mu_n(x)-\mu(x)|$, where $T=[l,u]$. With the latter result, one can construct an asymptotic $(1-\alpha)$ SCB, $0<\alpha<1$, by finding two functions $\mu_n^{\mathrm{lower}}(x)$ and $\mu_n^{\mathrm{upper}}(x)$, such that

$$\lim_{n\to\infty} \mathsf{P}(\mu_n^{\mathrm{lower}}(x)\le\mu(x)\le\mu_n^{\mathrm{upper}}(x) \text{ for all } x\in T)=1-\alpha. \tag{1.6}$$

The SCB can be used for model validation: one can test whether $\mu(\cdot)$ is of certain parametric functional form by checking whether the fitted parametric form lies in the SCB. Following the work of Bickel and Rosenblatt (1973), Johnston (1982) derived the asymptotic distribution of $\sup_{0\le x\le 1}|\mu_n(x)-\mathsf{E}[\mu_n(x)]|$, assuming that (X_i,Y_i) are independent random samples from a bivariate population. Johnston's derivation is no longer valid if dependence is present. For other work on regression confidence bands under independence see Knafl, Sacks and Ylvisaker (1985), Hall and Titterington (1988), Härdle and Marron (1991), Sun and Loader (1994), Xia (1998), Cummins, Filloon and Nychka (2001) and Dümbgen (2003), among others. Recently Zhao and Wu (2008) proposed a method for constructing SCB for stochastic regression models which have asymptotically correct coverage probabilities. However, their confidence band is over an increasingly dense grid of points instead of over an interval [see also Bühlmann (1998) and Knafl, Sacks and Ylvisaker (1985)]. Here we shall also solve the latter problem and establish a uniform asymptotic theory for the regression estimate $\mu_n(x)$, so that one can construct a genuine SCB for regression functions. A similar result will be derived for $\sigma(\cdot)$ as well.

The rest of the paper is organized as follows. Main results are presented in Section 2. Proofs are given in Sections 4 and 5. Our results are applied in Section 3 to the U.S. Treasury yield rates data.

2. Main results. Before stating our theorems, we first introduce dependence measures. Assume $X_k \in \mathcal{L}^p$, $p > 0$. Here for a random variable W, we write $W \in \mathcal{L}^p$ $(p > 0)$, if $\|W\|_p := (\mathsf{E}|W|^p)^{1/p} < \infty$. Let $\{\varepsilon'_j\}_{j\in\mathsf{Z}}$ be an i.i.d. copy of $\{\varepsilon_j\}_{j\in\mathsf{Z}}$; let $\xi_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$ and

$$X'_n = G(\xi'_n) \qquad \text{where } \xi'_n = (\xi_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n).$$

Here X'_n is a coupled process of X_n with ε_0 in the latter replaced by an i.i.d. copy ε'_0. Following Wu (2005), define the physical dependence measure

$$\theta_{n,p} = \|X_n - X'_n\|_p.$$

Let $\theta_{n,p} = 0$ if $n < 0$. A similar quantity can be defined if we couple the whole past: let $\xi^\star_{k,n} = (\ldots, \varepsilon'_{k-n-2}, \varepsilon'_{k-n-1}, \xi_{k-n,k})$, $k \ge n$, where $\xi_{i,j} = (\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_j)$, and define

$$\Psi_{n,p} = \|G(\xi_n) - G(\xi^\star_{n,n})\|_p. \tag{2.1}$$

Our conditions on dependence will be expressed in terms of $\theta_{n,p}$ and $\Psi_{n,p}$.

2.1. *Kernel density estimates.* We first consider a special case of (1.2) in which X_n has the form

$$X_n = a_0\varepsilon_n + g(\ldots, \varepsilon_{n-2}, \varepsilon_{n-1}) = a_0\varepsilon_n + g(\xi_{n-1}), \tag{2.2}$$

where g is a measurable function and $a_0 \ne 0$. Then the coupled process $X'_n = a_0\varepsilon_n + g(\xi_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_{n-1})$. We need the following conditions:

(C1). There exists $0 < \delta_2 \le \delta_1 < 1$ such that $n^{-\delta_1} = O(b_n)$ and $b_n = O(n^{-\delta_2})$.

(C2). Suppose that $X_1 \in \mathcal{L}^p$ for some $p > 0$. Let $p' = \min(p, 2)$ and $\Theta_n = \sum_{i=0}^n \theta_{i,p'}^{p'/2}$. Assume $\Psi_{n,p'} = O(n^{-\gamma})$ for some $\gamma > \delta_1/(1-\delta_1)$ and

$$\mathcal{Z}_n b n^{-1} = o(\log n) \qquad \text{where } \mathcal{Z}_n = \sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2. \tag{2.3}$$

(C3). The density function f_ε of ε_1 is positive and

$$\sup_{x\in\mathsf{R}} [f_\varepsilon(x) + |f'_\varepsilon(x)| + |f''_\varepsilon(x)|] < \infty.$$

(C4). The support of K is $[-A, A]$, where K is differentiable over $(-A, A)$, the right (resp., left) derivative $K'(-A)$ [resp., $K'(A)$] exists, and $\sup_{|x|\le A} |K'(x)| < \infty$. The Lebesgue measure of the set $\{x \in [-A, A] : K(x) = 0\}$ is zero. Let $\lambda_K = \int K^2(y)\,dy$, $K_1 = [K^2(-A) + K^2(A)]/(2\lambda_K)$ and $K_2 = \int_{-A}^{A} (K'(t))^2\,dt/(2\lambda_K)$.

THEOREM 2.1. *Let $l, u \in \mathsf{R}$ be fixed and X_n be of form (2.2). Assume* (C1)–(C4). *Then we have for every $z \in \mathsf{R}$,*

$$\mathsf{P}((2\log \bar{b}^{-1})^{1/2}(\Delta_n - d_n) \le z) \to e^{-2e^{-z}}, \tag{2.4}$$

where $\bar{b} = b/(u-l)$,

$$d_n = (2\log \bar{b}^{-1})^{1/2} + \frac{1}{(2\log \bar{b}^{-1})^{1/2}}\left\{\log \frac{K_1}{\pi^{1/2}} + \frac{1}{2}\log\log \bar{b}^{-1}\right\},$$

if $K_1 > 0$, and otherwise

$$d_n = (2\log \bar{b}^{-1})^{1/2} + \frac{1}{(2\log \bar{b}^{-1})^{1/2}}\log \frac{K_2^{1/2}}{2^{1/2}\pi}.$$

We now discuss conditions (C1)–(C4). The bandwidth condition (C1) is fairly mild. In (C2), the quantity Θ_n measures the cumulative dependence of $X_0, \ldots, X_n$ on ε_0, and, with (C1), it gives sufficient dependence and bandwidth conditions for the asymptotic Gumbel convergence (2.4). For short-range dependent linear process $X_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j}$ with $\mathsf{E}\varepsilon_1 = 0$ and $\mathsf{E}\varepsilon_1^2 = 1$, (C2) is satisfied if $\sum_{j=0}^{\infty} |a_j| < \infty$ and $\sum_{j=n}^{\infty} a_j^2 = O(n^{-\gamma})$ for some $\gamma > 2\delta_1/(1-\delta_1)$. The latter condition can be weaker than $\sum_{j=0}^{\infty} |a_j| < \infty$ if $\delta_1 < 1/3$. Interestingly, (C2) also holds for some long-range dependent processes; see Theorem 2.3. With (C3), it is easily seen that X_i does have a density. If (C3) is violated, then X_i may not have a density. For example, if ε_i are i.i.d. Bernoulli with $\mathsf{P}(\varepsilon_i = 0) = \mathsf{P}(\varepsilon_i = 1) = 1/2$, then $X_0 = \sum_{i=0}^{\infty} \rho^i \varepsilon_{-i}$, where $\rho = (\sqrt{5} - 1)/2$, does not have a density [Erdös (1939)]. The kernel condition (C4) is quite mild and it is satisfied by many popular kernels. For example, it holds for the Epanechnikov kernel $K(u) = 0.75(1-u^2)\mathbf{1}_{|u| \le 1}$.

In Theorem 2.2 below, we do not assume the special form (2.2). We need regularity conditions on conditional density functions. For jointly distributed random vectors ξ and η, let $F_{\eta|\xi}(\cdot)$ be the conditional distribution function of η given ξ; let $f_{\eta|\xi}(x) = \partial F_{\eta|\xi}(x)/\partial x$ be the conditional density. For function g with $\mathsf{E}|g(\eta)| < \infty$, let $\mathsf{E}(g(\eta)|\xi) = \int g(x)\, dF_{\eta|\xi}(x)$ be the conditional expectation of $g(\eta)$ given ξ.

Conditions (C2) and (C3) are replaced, respectively, by:

(C2)$'$. Suppose that $X_1 \in \mathcal{L}^p$ and $\theta_{n,p} = O(\rho^n)$ for some $p > 0$ and $0 < \rho < 1$.

(C3)$'$. The density function f is positive and there exists a constant $B < \infty$ such that

$$\sup_x [|f_{X_n|\xi_{n-1}}(x)| + |f'_{X_n|\xi_{n-1}}(x)| + |f''_{X_n|\xi_{n-1}}(x)|] \le B \qquad \text{almost surely.}$$

THEOREM 2.2. *Under* (C1), (C2)′, (C3)′ *and* (C4), *we have (2.4).*

Many nonlinear time series models (e.g., ARCH models, bilinear models, exponential AR models) satisfy (C2)′; see Shao and Wu (2007). If (X_i) is a Markov chain of the form $X_i = R(X_{i-1}, \varepsilon_i)$, where $R(\cdot,\cdot)$ is a bivariate measurable function, then $f_{X_i|\xi_{i-1}}(\cdot)$ is the conditional density of X_i given X_{i-1}. Consider the ARCH model $X_i = \varepsilon_i(a^2 + b^2 X_{i-1}^2)^{1/2}$, where $a > 0, b > 0$ are real parameters and ε_i has density function f_ε, then $f_{X_i|X_{i-1}}(x) = f_\varepsilon(x/H_i)/H_i$, where $H_i = (a^2 + b^2 X_{i-1}^2)^{1/2}$. So (C3)′ holds if $\sup_x[f_\varepsilon(x) + |f'_\varepsilon(x)| + |f''_\varepsilon(x)|] < \infty$ [cf. (C3)]. For more general ARCH-type processes see Doukhan, Madre and Rosenbaum (2007).

For short-range dependent processes for which

$$\Theta_\infty = \sum_{i=0}^{\infty} \theta_{i,p'}^{p'/2} < \infty, \tag{2.5}$$

we have $\mathcal{Z}_n = O(n)$ and (2.3) of condition (C2) trivially holds. For long-range dependent processes, (2.5) can be violated. A popular model for long-range dependence is the fractionally integrated auto-regressive moving average process [Granger and Joyeux (1980), Hosking (1981)]. Here we consider the more general form of linear processes with slowly decaying coefficients:

$$X_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j} \qquad \text{where } a_j = j^{-\beta}\ell(j), 1/2 < \beta < 1. \tag{2.6}$$

Here $a_0 = 1$, $\ell(\cdot)$ is a slowly varying function and ε_i are i.i.d. with $\mathsf{E}\varepsilon_i = 0$ and $\mathsf{E}\varepsilon_i^2 = 1$.

THEOREM 2.3. *Assume (2.6). Let $l, u \in \mathsf{R}$ be fixed.* (i) *Assume* (C1), (C3), (C4), $\delta_1/(1-\delta_1) < \beta - 1/2$ *and*

$$b_n^{1/2} n^{1-\beta} \ell(n) = o(\log^{-1/2} n). \tag{2.7}$$

Then (2.4) holds. (ii) *Assume* (C1), (C3), (C4), $\sup_x |f'''_\varepsilon(x)| < \infty$ *and*

$$\log^{1/2} n = o(b_n^{1/2} n^{1-\beta} \ell(n)). \tag{2.8}$$

Let $c_\beta = \int_0^\infty (x + x^2)^{-\beta}\, dx/[(3-2\beta)(1-\beta)]$. *Then*

$$\frac{\Delta_n}{b_n^{1/2} n^{1-\beta} \ell(n)} \Rightarrow |N(0,1)| \frac{\sqrt{c_\beta}}{\sqrt{\lambda_K}} \max_{l \le x \le u} \frac{|f'(x)|}{\sqrt{f(x)}}. \tag{2.9}$$

Theorem 2.3 reveals the interesting dichotomy phenomenon for the maximum deviation Δ_n: if the bandwidth b_n is small such that (2.7) holds, then

the asymptotic distribution is the same as the one under short-range dependence. However, if b_n is large, then both the normalizing constant and the asymptotic distribution change. Let $b_n = n^{-\delta}\ell_1(n)$, where ℓ_1 is another slowly varying function. Simple algebra shows that, if $\max((1+\delta)/(1-\delta), 2-\delta) < 2\beta$, then the bandwidth condition in Theorem 2.3(i) holds. The latter inequality requires $\beta > \sqrt{3}/2 = 0.866025, \ldots$. If $\beta < 1-\delta/2$, then (2.8) holds. Theorem 2.3(ii) is similar to Theorem 3.1 in Ho and Hsing (1996), with our result having a wider range of β.

2.2. *Estimation of $\mu(\cdot)$ and $\sigma^2(\cdot)$.* Let $\widetilde{\xi}_i = (\ldots, \eta_{i-1}, \eta_i, \xi_i)$. For a function h with $\mathsf{E}h^2(\eta_i) < \infty$, write

$$M_n^r(x) = \frac{1}{nb}\sum_{k=1}^{n} K\left(\frac{X_k - x}{b}\right) Z_k \qquad \text{where } Z_k = h(\eta_k) - \mathsf{E}h(\eta_k).$$

PROPOSITION 2.1. *Let $l, u \in \mathsf{R}$ be fixed. Assume $\sigma^2 = \mathsf{E}Z_1^2$ and $\mathsf{E}|Z_1|^p < \infty$, $p > 2/(1-\delta_1)$.* (i) *Assume (2.2),* (C1), (C3)–(C4) *and $\Psi_{n,q} = O(n^{-\gamma})$ for some $q > 0$ and $\gamma > \delta_1/(1-\delta_1)$. Then for all $z \in \mathsf{R}$,*

$$\mathsf{P}\left(\sqrt{\frac{nb}{\lambda_K}} \sup_{l \le x \le u} \frac{|M_n^r(x)|}{f^{1/2}(x)\sigma} - d_n \le \frac{z}{(2\log \bar{b}^{-1})^{1/2}}\right) \to e^{-2e^{-z}} \tag{2.10}$$

as $n \to \infty$. (ii) *Assume (1.2),* (C1), (C2)$'$, (C3)$'$ *and* (C4) *hold with ξ_{n-1} in* (C2)$'$ *replaced by $\widetilde{\xi}_{n-1}$. Then (2.10) holds.*

Proposition 2.1(i) allows for long-range dependent processes. For (2.6), by Karamata's theorem, $\Psi_{n,2} = O(n^{1/2-\beta}\ell(n))$. So we have $\Psi_{n,2} = O(n^{-\gamma})$ with $\gamma > \delta_1/(1-\delta_1)$ if $\delta_1 < (2\beta-1)/(2\beta+1)$.

For $S \subset \mathsf{R}$, denote by $\mathcal{C}^p(S) = \{g(\cdot) : \sup_{x \in S} |g^{(k)}(x)| < \infty, k = 0, \ldots, p\}$ the set of functions having bounded derivatives on S up to order $p \ge 1$. Let $S^\epsilon = \bigcup_{y \in S}\{x : |x-y| \le \epsilon\}$ be the ϵ-neighborhood of S, $\epsilon > 0$.

THEOREM 2.4. *Let $l, u \in \mathsf{R}$ be fixed and K be symmetric. Assume that the conditions in Proposition 2.1 hold with $Z_n = \eta_n$, $f_\varepsilon(\cdot), \mu(\cdot) \in \mathcal{C}^4(T^\epsilon)$ for some $\epsilon > 0$, where $T = [l, u]$, and that b satisfies*

$$0 < \delta_1 < 1/3, \qquad nb^9 \log n = o(1) \quad \textit{and} \quad \mathcal{Z}_n b^3 = o(n \log n).$$

Let $\psi_K = \int u^2 K(u)\, du/2$ and $\rho_\mu(x) = \mu''(x) + 2\mu'(x)f'(x)/f(x)$. Then

$$\mathsf{P}\left(\sqrt{\frac{nb}{\lambda_K}} \sup_{l \le x \le u} \frac{\sqrt{f_n(x)}|\mu_n(x) - \mu(x) - b^2\psi_K\rho_\mu(x)|}{\sigma(x)} - d_n \le \frac{z}{(2\log \bar{b}^{-1})^{1/2}}\right) \to e^{-2e^{-z}}. \tag{2.11}$$

Note that $\sigma^2(x) = \mathsf{E}[(Y_k - \mu(X_k))^2 | X_k = x]$. It is natural to use the Nadaraya–Watson method to estimate $\sigma^2(x)$ based on the residuals $\hat{e}_k = Y_k - \mu_n(X_k)$:

$$\sigma_n^2(x) = \frac{1}{nhf_{n1}(x)} \sum_{k=1}^n K\left(\frac{X_k - x}{h}\right)[Y_k - \mu_n(X_k)]^2,$$

where the bandwidths $h = h_n \to 0$ and $nh_n \to \infty$, and

$$f_{n1}(x) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{X_k - x}{h}\right).$$

THEOREM 2.5. *Let $l, u \in \mathsf{R}$ be fixed and K be symmetric. Assume $\nu_\eta = \mathsf{E}\eta_1^4 - 1 < \infty$. Further assume that the conditions in Proposition 2.1 hold with $Z_n = \eta_n^2 - 1$, $f(\cdot), \sigma(\cdot) \in \mathcal{C}^4(T^\epsilon)$ for some $\epsilon > 0$, where $T = [l, u]$, and that $h \asymp b$ satisfies*

$$0 < \delta_1 < 1/4, \qquad nb^9 \log n = o(1)$$

and

$$\mathcal{Z}_n b^3 = o(n \log n).$$

Let $\rho_\sigma(x) = 2\sigma'^2(x) + 2\sigma(x)\sigma''(x) + 4\sigma(x)\sigma'(x)f'(x)/f(x)$. Then

$$\mathsf{P}\left(\sqrt{\frac{nh}{\lambda_K \nu_\eta}} \sup_{l \le x \le u} \frac{\sqrt{f_{n1}(x)}|\sigma_n^2(x) - \sigma^2(x) - h^2 \psi_K \rho_\sigma(x)|}{\sigma^2(x)} - d_n \le \frac{z}{(2 \log \bar{h}^{-1})^{1/2}}\right) \to e^{-2e^{-z}}, \tag{2.12}$$

where d_n is defined as in Theorem 2.1 by replacing $\bar{b}$ with $\bar{h} = h/(u - l)$.

We now compare the SCBs constructed based on Theorem 1 in Zhao and Wu (2008) and Theorem 2.4. Assume $l = 0$ and $u = 1$. The former is over the grid point $T_n = \{2b_n j, j = 0, 1, \ldots, J_n\}$ with $J_n = \lceil 1/(2b_n) \rceil$, while the latter is a genuine SCB in the sense that it is over the whole interval $T = [0, 1]$. Let $\hat{\rho}_\mu(\cdot)$ [resp., $\hat{\sigma}(\cdot)$] be a consistent estimate of $\rho_\mu(\cdot)$ [resp., $\sigma(\cdot)$] and $z_\alpha = -\log\log(1 - \alpha)^{-1/2}$, $0 < \alpha < 1$. By Theorem 2.4, we can construct the $1 - \alpha$ SCB for $\mu(x)$ over $x \in [0, 1]$ as

$$\mu_n(x) - b^2 \psi_K \hat{\rho}_\mu(x) \pm l_1 \hat{\sigma}(x) \sqrt{\frac{\lambda_K}{nbf_n(x)}} \quad \text{where } l_1 = \frac{z_\alpha}{(2 \log b^{-1})^{1/2}} + d_n. \tag{2.13}$$

Similarly, using Theorem 1 in Zhao and Wu (2008), the $1-\alpha$ confidence band for $\mu(x)$ over $x \in T_n$ is also of form (2.13) with l_1 replaced by

$$l_2 = \frac{z_\alpha}{(2\log J_n)^{1/2}} + (2\log J_n)^{1/2} - \frac{1/2\log\log J_n + \log(2\sqrt{\pi})}{(2\log J_n)^{1/2}}.$$

Elementary calculations show that, interestingly, l_1 and l_2 are quite close: $l_1 - l_2 = (\log\log b^{-1})/(2\log b^{-1})^{1/2}(1+o(1))$ if $K_1 > 0$.

3. Application to the treasury bill data. There is a huge literature on models for short-term interest rates. Let R_t be the interest rate at time t. Assume that R_t follows the diffusion model

$$dR_t = \mu(R_t)\,dt + \sigma(R_t)\,d\mathbb{B}(t), \tag{3.1}$$

where $\mathbb{B}$ is the standard Brownian motion, $\mu(\cdot)$ is the instantaneous return or drift function and $\sigma(\cdot)$ is the volatility function. Black and Scholes (1973) considered the model with $\mu(x) = \alpha x$ and $\sigma(x) = \sigma x$. Vasicek (1977) assumed that $\mu(x) = \alpha_0 + \alpha_1 x$ and $\sigma(x) \equiv \sigma$, where α_0, α_1 and σ are unknown constants. Cox, Ingersoll and Ross (1985) and Courtadon (1982) assumed that $\sigma(x) = \sigma x^{1/2}$ and $\sigma(x) = \sigma x$, respectively. Both models are generalized by Chan et al. (1992) to the form $\sigma(x) = \sigma x^\gamma$, with σ and γ being unknown parameters. Stanton (1997), Fan and Yao (1998), Chapman and Pearson (2000) and Fan and Zhang (2003) considered the nonparametric estimation of $\mu(\cdot)$ and $\sigma(\cdot)$ in (3.1); see also Aït-Sahalia (1996a, 1996b). Stanton (1997) constructed *point-wise* confidence intervals which serve as a tool for suggesting which parametric models to use. Zhao (2008) gave an excellent review of parametric and nonparametric approaches of (3.1). See also the latter paper for further references.

Here we shall consider the U.S. six-month treasury yield rates data from January 2nd, 1990 to July 31st, 2009. The data can be downloaded from the U.S. Treasury department's website `http://www.ustreas.gov/`. It has 4900 daily rates and a plot is given in Figure 1. Let $X_i = R_{t_i}$ be the rate at day $i = 1, \dots, 4900$. For the daily data, since one year has 250 transaction days, $t_i - t_{i-1} = 1/250$. Let $\Delta = 1/250$. As a discretized version of (3.1), we consider the model

$$Y_i = \mu(X_i)\Delta + \sigma(X_i)\Delta^{1/2}\eta_i, \tag{3.2}$$

where $Y_i = R_{t_{i+1}} - R_{t_i} = X_{i+1} - X_i$ and $\eta_i = (\mathbb{B}(t_{i+1}) - \mathbb{B}(t_i))/\Delta^{1/2}$ are i.i.d. standard normal. For convenience of applying Theorem 2.4, in the sequel we shall write $\mu(X_i)\Delta$ [resp., $\sigma(X_i)\Delta^{1/2}$] in (3.2) as $\mu(X_i)$ [resp., $\sigma(X_i)$]. So (3.2) is rewritten as

$$Y_i = \mu(X_i) + \sigma(X_i)\eta_i. \tag{3.3}$$

Figure 2 shows the estimated 95% simultaneous confidence band for the regression function $\mu(\cdot)$ over the interval $T = [l, u] = [0.35, 8.06]$, which includes 96% of the daily rates X_i. To select the bandwidth, we use the `R` program `bw.nrd` which gives $b = 0.37$. Then we use the `R` program `locpoly` for local polynomial regression. The Nadaraya–Watson estimate is a special case of the local polynomial regression with degree 0. The function $\rho(x)$ in the bias term $b^2\psi_K\rho(x)$ in Theorem 2.4 involves the first and second order derivatives μ', f' and μ''. The program `locpoly` can also be used to estimate derivatives μ' and μ'', where we use the bigger bandwidth $2b = 0.74$. For f, we use the `R` program `density`, and estimate f' by differentiating the estimated density. Then we can have the bias-corrected estimate $\tilde{\mu}_n(x) = \mu_n(x) - b^2\psi_K\hat{\rho}(x)$ for μ, which is plotted in the the middle curve in Figure 2. To estimate $\sigma(\cdot)$, as in Stanton (1997), we shall make use of the estimated residuals $\hat{e}_i = Y_i - \tilde{\mu}_n(X_i)$, and perform the Nadaraya–Watson regression of $\hat{e}_i^2$ versus X_i with the bandwidth b. In our data analysis the boundary problem of the Nadaraya–Watson regression raised in Chapman and Pearson (2000) is not severe since we focus on the interval $T = [0.35, 8.06]$, while the whole range is $[\min X_i, \max X_i] = [0.14, 8.49]$.

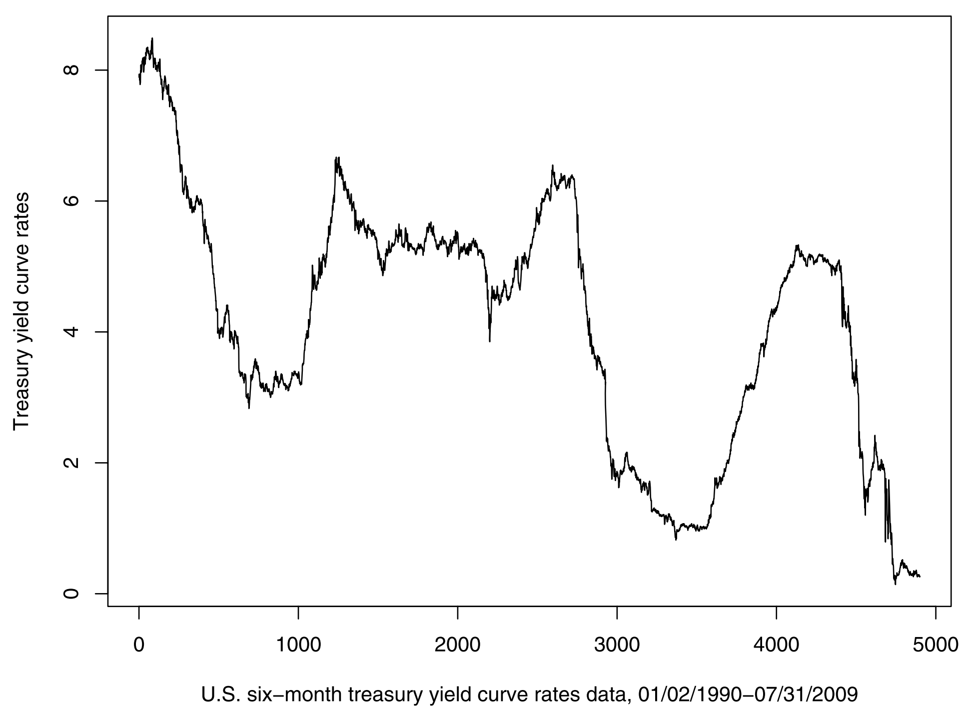

FIG. 1. *U.S. six-month treasury yield curve rates data from January 2nd, 1990 to July 31st, 2009. Source: U.S. Treasury department's website* http://www.ustreas.gov/.

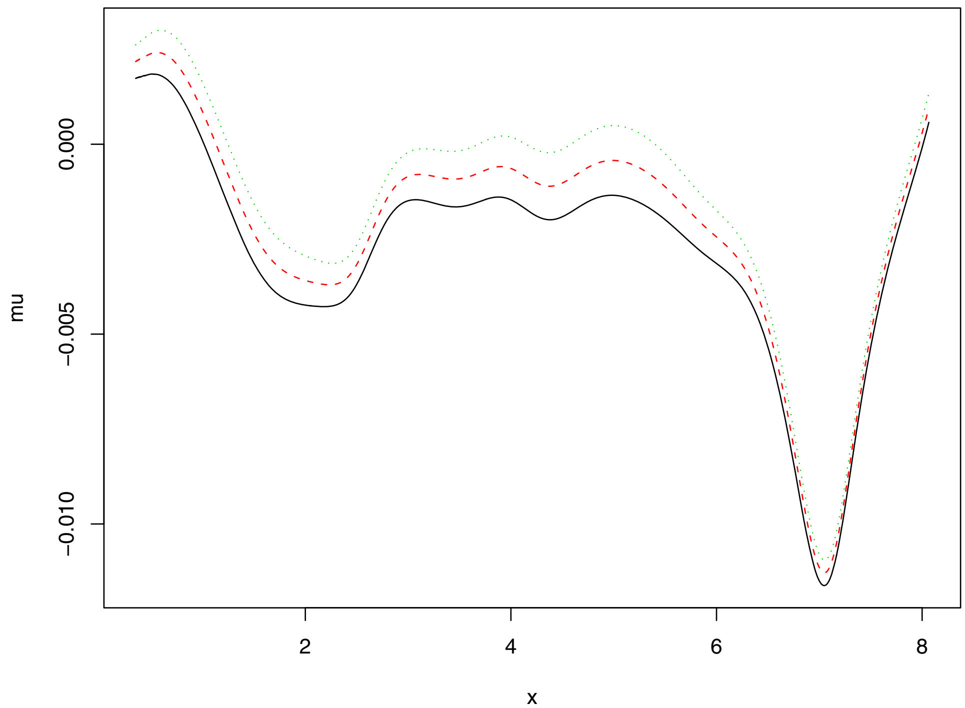

FIG. 2. 95% *SCB of the regression function* $\mu(\cdot)$ *over the interval* $[l,u]=[0.35,8.06]$. *The dashed curve in the middle is* $\mu_n(x)-b^2\psi_K\hat{\rho}(x)$, *the bias-corrected estimate of* μ.

The Gumbel convergence in Theorem 2.4 can be quite slow, so the SCB in (2.13) may not have a good finite-sample performance. To circumvent this problem, we shall adopt a simulation based method. Let

$$\Pi_n=\sup_{x\in T}\frac{|\sum_{k=1}^n K(X_k^*/b-x/b)\eta_k^*|}{nbf^{1/2}(x)},$$

where X_k^* are i.i.d. with density f, η_k^* are i.i.d. with $\mathsf{E}\eta_n=0$, $\mathsf{E}\eta_n^2=1$ and $\mathsf{E}|\eta_1|^p<\infty$, and (X_k^*) and (η_k^*) are independent. As in Theorem 2.4, let

$$\Pi_n'=\sup_{x\in T}\frac{\sqrt{f(x)}|\mu_n(x)-\mu(x)-b^2\psi_K\rho(x)|}{\sigma(x)}.$$

By Theorem 2.4 and Proposition 2.1, with proper centering and scaling, Π_n and Π_n' have the same asymptotic Gumbel distribution. So the cutoff value, the $(1-\alpha)$th quantile of Π_n', can be estimated by the sample $(1-\alpha)$th quantile of many simulated Π_n's. For the U.S. Treasury bill data, we simulated 10,000 Π_n's and obtained the 95% sample quantile 0.39. Then the SCB is constructed as $\tilde{\mu}_n(x)\pm 0.39\hat{\sigma}(x)/f_n^{1/2}(x)$; see the upper and lower curves in Figure 2.

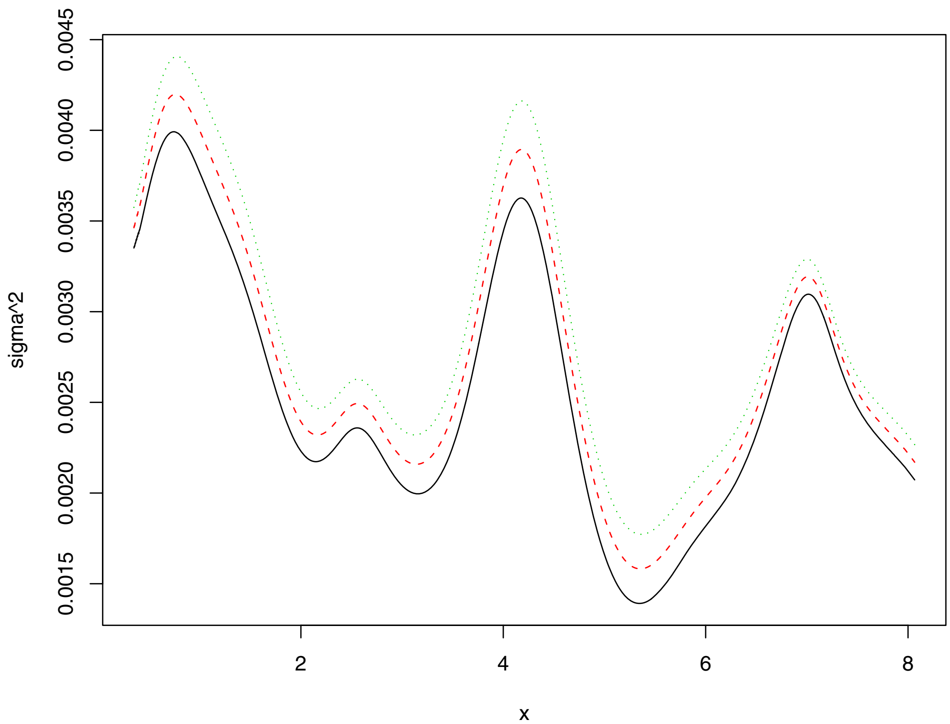

FIG. 3. 95% *SCB of the volatility function* $\sigma^2(\cdot)$ *over the interval* $[l, u] = [0.35, 8.06]$. *The dashed curve in the middle is* $\sigma_n^2(x) - b^2\psi_K\hat{\rho}_\sigma(x)$, *the bias-corrected estimate of* σ^2.

We now apply Theorem 2.5 to construct SCB for $\sigma^2(\cdot)$. We choose $h = b$, which has a reasonably satisfactory performance in our data analysis. By Theorem 2.5,

$$\Pi_n'' = \frac{1}{\sqrt{\nu_\eta}} \sup_{x \in T} \frac{\sqrt{f(x)}|\sigma_n^2(x) - \sigma^2(x) - b^2\psi_K\rho_\sigma(x)|}{\sigma^2(x)}$$

has the same asymptotic distribution as Π_n and Π_n'. Based on the above simulation, we choose the cutoff value 0.39. As in the treatment of μ' and μ'' in the bias term of μ_n, we use a similar estimate, noting that $\rho_\sigma(x) = (\sigma^2(x))'' + 2(\sigma^2(x))' f'(x)/f(x)$ has the same form as $\rho_\mu(x)$. The 95% SCB of $\sigma^2(\cdot)$ is presented in Figure 3.

Based on the 95% SCB of $\mu(\cdot)$, we conclude that the linear drift function hypothesis $H_0 : \mu(x) = \alpha_0 + \alpha_1 x$ for some α_0 and α_1 is rejected at the 5% level. Other simple parametric forms do not seem to exist. Similar claims can be made for $\sigma^2(\cdot)$, and none of the parametric forms previously mentioned seems appropriate. This suggests that the dynamics of the treasury yield rates might be far more complicated than previously speculated.

4. Proofs of Theorems 2.1–2.3. Throughout the proofs C denotes constants which do not depend on n and b_n. The values of C may vary from place to place. Let $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ be the floor and ceiling functions, respectively. Without loss of generality, we assume $l=0$, $u=1$ in (1.5) and $A=1$ in condition (C4). Write

$$\frac{\sqrt{nb}}{\sqrt{\lambda_K f(bt)}}[f_n(bt)-\mathsf{E}f_n(bt)]=M_n(t)+N_n(t),$$

where $M_n(t)$ has summands of martingale differences

$$M_n(t)=\frac{1}{\sqrt{nb\lambda_K f(bt)}}\sum_{k=1}^{n}\{K(X_k/b-t)-\mathsf{E}[K(X_k/b-t)|\xi_{k-1}]\},$$

and, since $\mathsf{E}[K(X_k/b-t)|\xi_{k-1}]=b\int_{-1}^{1}K(v)f_{X_k|\xi_{k-1}}(bv+bt)\,dv$, the remainder

$$\begin{aligned}N_n(t)&=\frac{1}{\sqrt{nb\lambda_K f(bt)}}\sum_{k=1}^{n}\{\mathsf{E}[K(X_k/b-t)|\xi_{k-1}]-\mathsf{E}K(X_k/b-t)\}\\&=\frac{\sqrt{b}}{\sqrt{n\lambda_K f(bt)}}\int_{-1}^{1}K(v)Q_n'(bv+bt)\,dv,\end{aligned}$$

where

$$Q_n(x)=\sum_{k=1}^{n}[F_{X_k|\xi_{k-1}}(x)-F(x)].$$

If X_n admits the form (2.2), we assume $a_0=1$. Let $Y_k=g(\ldots,\varepsilon_{k-1},\varepsilon_k)$. Then $f_{X_k|\xi_{k-1}}(bv+bt)=f_\varepsilon(bv+bt-Y_{k-1})$.

Proofs of Theorems 2.1 and 2.2. We split $[1,n]$ into alternating big and small blocks H_1, $I_1,\ldots,H_{\iota_n}$, I_{ι_n}, I_{ι_n+1}, with length $|H_i|=\lfloor n^{\tau_1}\rfloor$, $|I_i|=\lfloor n^{\tau}\rfloor$, $1\le i\le\iota_n$, $|I_{\iota_n+1}|=n-\iota_n(\lfloor n^{\tau_1}\rfloor+\lfloor n^{\tau}\rfloor)$ and $\iota_n=\lfloor n/(\lfloor n^{\tau_1}\rfloor+\lfloor n^{\tau}\rfloor)\rfloor$, where $\delta_1/\gamma<\tau<\tau_1<1-\delta_1$. Let $m=|I_1|$,

$$u_j(t)=\sum_{k\in H_j}\{\mathsf{E}[K(X_k/b-t)|\xi_{k-m,k}]-\mathsf{E}[K(X_k/b-t)|\xi_{k-m,k-1}]\},$$

$$v_j(t)=\sum_{k\in I_j}\{\mathsf{E}[K(X_k/b-t)|\xi_{k-m,k}]-\mathsf{E}[K(X_k/b-t)|\xi_{k-m,k-1}]\},$$

$$\widetilde{M}_n(t)=\frac{1}{\sqrt{nb\lambda_K f(bt)}}\sum_{j=1}^{\iota_n}u_j(t),\qquad R_n(t)=\frac{1}{\sqrt{nb\lambda_K f(bt)}}\sum_{j=1}^{\iota_n+1}v_j(t).$$

Theorems 2.1 and 2.2 follow from Lemmas 4.1–4.3 and Lemma 4.5 below. □

PROOF OF THEOREM 2.3. Case (i) follows from Theorem 2.1. For (ii), since $\sum_{i=1}^{n} Y_{i-1}/(c_\beta n^{3/2-\beta}\ell(n)) \Rightarrow N(0,1)$ [cf. Ho and Hsing (1996)], where $Y_{i-1} = \sum_{k=1}^{\infty} a_k \varepsilon_{i-k}$, it follows from (2.8), Lemma 4.1(ii) and Lemma 4.4. □

LEMMA 4.1. *Assume* (C4). (i) *We have*

$$\sup_{0\le t\le b^{-1}} |N_n(t)| = O_{\mathsf{P}}(b^{1/2} n^{-1/2} \widetilde{\Theta}_n), \tag{4.1}$$

where $\widetilde{\Theta}_n = \mathcal{Z}_n^{1/2}$ *if* (X_n) *satisfies (2.2) and* (C3); $\widetilde{\Theta}_n = O(n^{1/2})$ *if* (X_n) *satisfies (1.2),* (C2)′ *and* (C3)′. (ii) *For the process (2.6), we have (4.1) with* $\widetilde{\Theta}_n = O(n^{3/2-\beta}\ell(n))$, *and*

$$\sup_{0\le t\le b^{-1}} \left| N_n(t)\sqrt{nb\lambda_K f(bt)} - bf'(bt)\sum_{j=1}^{n} Y_{j-1} \right| = o(bn^{3/2-\beta}\ell(n)), \tag{4.2}$$

where $Y_{j-1} = \sum_{k=1}^{\infty} a_k \varepsilon_{j-k}$.

LEMMA 4.2. *Under conditions of Theorems 2.1 or 2.2, we have*

$$\mathsf{P}\Big(\sup_{0\le t\le b^{-1}} |M_n(t) - \widetilde{M}_n(t) - R_n(t)| \ge (\log b^{-1})^{-2}\Big) = o(1).$$

LEMMA 4.3. *Under conditions of Theorems 2.1 or 2.2, we have*

$$\mathsf{P}\Big(\sup_{0\le t\le b^{-1}} |R_n(t)| \ge (\log b^{-1})^{-2}\Big) = o(1). \tag{4.3}$$

LEMMA 4.4. *Let* $\sup_x f_{X_n|\xi_{n-1}}(x)$ *be a.s. bounded. Assume* (C4). *Then*

$$\sup_{0\le t\le b^{-1}} |M_n(t)| = O_{\mathsf{P}}(\sqrt{\log n}).$$

Consequently, under conditions of Lemma 4.1, $\mathsf{E}f_n(x) - f(x) = f''(x)b^2\psi_K + o(b^2)$ *and*

$$\sup_{0\le x\le 1} |f_n(x) - f(x)| = \frac{O_{\mathsf{P}}(\sqrt{\log n})}{\sqrt{nb}} + \frac{O_{\mathsf{P}}(\tilde{\Theta}_n)}{n} + O(b^2).$$

Lemma 4.4 gives an upper bound of $\sup_{0\le t\le b^{-1}} |M_n(t)|$. Under stronger conditions, one can have a far deeper asymptotic distributional result. By Lemmas 4.5, 4.2 and 4.3, it is asymptotically distributed as Gumbel.

LEMMA 4.5. *Under conditions of Theorems 2.1 or 2.2, we have for all* $z \in \mathsf{R}$ *that*

$$\mathsf{P}\Big(\sup_{0\le t\le b^{-1}} |\widetilde{M}_n(t)| < x_z\Big) \to e^{-2e^{-z}} \qquad \textit{where } x_z = d_n + \frac{z}{(2\log b^{-1})^{1/2}}. \tag{4.4}$$

4.1. *Proofs of Lemmas 4.1–4.4.*

PROOF OF LEMMA 4.1. We claim that, for any $a_0 > 0$,

$$\mathsf{E}\Big[\sup_{|x|\le a_0} |Q_n'(x)|^2\Big] = O(\widetilde{\Theta}_n^2), \tag{4.5}$$

which implies Lemma 4.1(i) in view of

$$N_n(t) = \frac{\sqrt{b}}{\sqrt{n\lambda_K f(bt)}} \int_{-1}^{1} K(x) Q_n'(b(x+t))\,dx \tag{4.6}$$

by noting that $\inf_{0\le x\le 1} f(x) > 0$, $\int_{-1}^{1} |K(u)|\,du < \infty$. To prove (4.5), we use Lemma 4 in Wu (2003), which implies that

$$\sup_{|x|\le a_0} |Q_n'(x)|^2 \le 2a_0^{-1} \int_{-a_0}^{a_0} |Q_n'(x)|^2\,dx + 2a_0 \int_{-a_0}^{a_0} |Q_n''(x)|^2\,dx.$$

We first suppose that (X_n) satisfies (2.2) and (C3). Let

$$\mathcal{P}_k \cdot = \mathsf{E}(\cdot|\mathcal{F}_k) - \mathsf{E}(\cdot|\mathcal{F}_{k-1}), \qquad k \in \mathsf{Z},$$

be the projection operators. By the orthogonality of $\mathcal{P}_k$, we have

$$\begin{aligned}\|Q_n'(x)\|_2^2 &= \sum_{k=-\infty}^{n} \|\mathcal{P}_k Q_n'(x)\|_2^2 \le \sum_{k=-\infty}^{n} \left(\sum_{i=1}^{n} \|\mathcal{P}_k f_{X_i|\xi_{i-1}}(x)\|_2\right)^2 \\ &\le C \sum_{k=-\infty}^{n} \left(\sum_{i=1-k}^{n-k} \theta_{i,p'}^{p'/2}\right)^2 = C\mathcal{Z}_n,\end{aligned}$$

where C does not depend on x. Similarly, we have $\sup_{x\in\mathsf{R}} \|Q_n''(x)\|_2^2 \le C\mathcal{Z}_n$. This proves (4.5).

To prove (4.5) for (X_n) satisfying (1.2), (C2)$'$ and (C3)$'$, we note that

$$\begin{aligned}\sup_{x\in\mathsf{R}} \|\mathcal{P}_k F_{X_i|\xi_{i-1}}(x)\|_2^2 &\le \sup_{x\in\mathsf{R}} \mathsf{E}|I\{X_i \le x\} - I\{X_{i,\{k\}} \le x\}| \\ &\le \sup_{x\in\mathsf{R}} \mathsf{P}(|X_i - x| \le |X_i - X_{i,\{k\}}|) \\ &\le C(\theta_{i-k,p}^{1/2} + \theta_{i-k,p}^{p/2}),\end{aligned}$$

where $X_{i,\{k\}} = G(\xi_{k-1}, \varepsilon'_k, \xi_{k+1,i})$ and we used the inequality

$$|I\{X \le x\} - I\{Y \le x\}| \le I\{|X - x| \le |X - Y|\}.$$

Since $\sup_x |f'_{X_n|\xi_{n-1}}(x)| \le B$, we have

$$\left| f_{X_i|\xi_{i-1}}(x) - \frac{F_{X_i|\xi_{i-1}}(x) - F_{X_i|\xi_{i-1}}(x-\Delta)}{\Delta} \right| \le B\Delta,$$

which by letting $\Delta = (\theta_{i-k,p}^{1/2} + \theta_{i-k,p}^{p/2})^{1/2}$ yields that

$$\sup_{x\in\mathsf{R}} \|\mathcal{P}_k f_{X_i|\xi_{i-1}}(x)\|_2^2 \le C(\theta_{i-k,p}^{1/2} + \theta_{i-k,p}^{p/2})^{1/2}.$$

This implies $\sup_{x\in\mathsf{R}} \|Q'_n(x)\|_2^2 = O(n)$. Similarly, we have $\sup_{x\in\mathsf{R}} \|Q''_n(x)\|_2^2 = O(n)$. We finish the proof of Lemma 4.1(i).

We now prove (4.2). For $i \ge 2$ write $Y_{i-1} = U + a_i\varepsilon_0 + W$, where $U = \sum_{j=1}^{i-1} a_j\varepsilon_{i-j}$ and $W = \sum_{j=i+1}^{\infty} a_j\varepsilon_{i-j}$. Let $W' = \sum_{j=i+1}^{\infty} a_j\varepsilon'_{i-j}$. Let $c_0 = \sup_x[|f'_\varepsilon(x)| + |f''_\varepsilon(x)|]$. By Taylor's expansion, there exists $R \in [0,1]$ such that

$$\begin{aligned}
\vartheta_i &:= \sup_x \|f_\varepsilon(x - Y_{i-1}) - f_\varepsilon(x - U - W) + a_i\varepsilon_0 f'_\varepsilon(x - U - a_i\varepsilon'_0 - W')\| \\
&= \sup_x \|-a_i\varepsilon_0 f'_\varepsilon(x - U - Ra_i\varepsilon_0 - W) + a_i\varepsilon_0 f'_\varepsilon(x - U - a_i\varepsilon'_0 - W')\| \\
&\le \|a_i\varepsilon_0 c_0 \min(1, |a_i\varepsilon'_0| + |a_i\varepsilon_0| + |W| + |W'|)\| = o(|a_i|).
\end{aligned}$$

Here we use the fact that $\|\varepsilon_0 \min(1, |a_i\varepsilon_0|)\| \to 0$ since $a_i \to 0$, and $a_i\varepsilon_0$ and $|W| + |W'|$ are independent. Since $\varepsilon'_l, \varepsilon_m$, $l, m \in \mathsf{Z}$, are i.i.d., we have $f(x) = \mathsf{E}[f_\varepsilon(x - U - a_i\varepsilon'_0 - W')|\xi_0]$. By the Lebesgue dominated convergence theorem, $f'(x) = \mathsf{E}[f'_\varepsilon(x - U - a_i\varepsilon'_0 - W')|\xi_0]$. By Jensen's inequality,

$$\sup_x \|\mathsf{E}[f_\varepsilon(x - Y_{i-1}) - f_\varepsilon(x - U - W)|\xi_0] + a_i\varepsilon_0 f'(x)\| \le \vartheta_i,$$

which again by Jensen's inequality implies that $\sup_x \|\mathsf{E}[f_\varepsilon(x - Y_{i-1}) - f_\varepsilon(x - U - W)|\xi_{-1}] \le \vartheta_i$. Since $\mathsf{E}[f_\varepsilon(x - U - W)|\xi_{-1}] = \mathsf{E}[f_\varepsilon(x - U - W)|\xi_0]$, we have

$$\sup_x \|\mathcal{P}_0[f_\varepsilon(x - Y_{i-1}) + f'(x)Y_{i-1}]\| \le 2\vartheta_i = o(|a_i|).$$

Define $\vartheta_i = 0$ if $i < 0$. Let $T_n(x) = Q_n(x) + f(x)\sum_{i=1}^n Y_{i-1}$. If $k \le -n$, then

$$\|\mathcal{P}_k T'_n(x)\| \le \sum_{j=1}^n 2\vartheta_{j-k} = o(n|k|^{-\beta}\ell(|k|)).$$

If $-n < k \le n$, by Karamata's theorem, $\sum_{i=1}^n a_i = O(na_n)$. Hence,

$$\sup_x \|\mathcal{P}_k T'_n(x)\| \le \sum_{j=1}^n 2\vartheta_{j-k} \le \sum_{j=1}^{2n} 2\vartheta_j = o(n^{1-\beta}\ell(n)).$$

Since $\mathcal{P}_k\cdot = \mathsf{E}(\cdot|\xi_k) - \mathsf{E}(\cdot|\xi_{k-1})$, $k \in \mathsf{Z}$, are orthogonal,

$$\sup_x \|T_n'(x)\|^2 = \sup_x \left(\sum_{k=-\infty}^{-n} + \sum_{k=1-n}^{n} \right) \|\mathcal{P}_k T_n'(x)\|^2 = o(n^{3-2\beta}\ell^2(n)),$$

where we again applied Karamata's theorem implying $\sum_{m=n}^{\infty} m^{-2\beta}\ell^2(m) = O(n^{1-2\beta}\ell^2(n))$. Similarly, since $\sup_x |f_\varepsilon'''(x)| < \infty$, we have $\sup_x \|T_n''(x)\|^2 = o(n^{3-2\beta}\ell^2(n))$. Since $T_n'(x) = T_n'(0) + \int_0^x T_n''(u)\,du$, for all finite $a_0 > 0$,

$$\mathsf{E}\Big[\sup_{|x|\le a_0} |T_n'(x)|^2\Big] = o(n^{3-2\beta}\ell^2(n)).$$

Hence, (4.2) follows in view of (4.6). $\square$

Proof of Lemma 4.2. Let $\widetilde{Z}_{k,t} = K(X_k/b - t) - \mathsf{E}[K(X_k/b-t)|\xi_{k-m,k}]$, $Z_{k,t} = \widetilde{Z}_{k,t} - \mathsf{E}(\widetilde{Z}_{k,t}|\xi_{k-1})$ and

$$[nb\lambda_K f(bt)]^{1/2}[M_n(t) - \widetilde{M}_n(t) - R_n(t)] = \sum_{k=1}^{n} Z_{k,t}.$$

We shall approximate $\sum_{k=1}^n Z_{k,t}$ by the skeleton process $\sum_{k=1}^n Z_{k,t_j}$, $1 \le j \le q_n$, where $q_n = \lfloor n^2/b \rfloor$ and $t_j = j/(bq_n)$. To this end, for $t \in [t_{j-1}, t_j]$, under condition (C4), if $X_k/b - t$ and $X_k/b - t_j$ are both in or outside $[-1,1]$, we have

$$|K(X_k/b - t) - K(X_k/b - t_j)| \le C|t - t_j| \le Cn^{-2}.$$

Otherwise, we have either $|X_k/b - t_j - 1| \le Cn^{-2}$ or $|X_k/b - t_j + 1| \le Cn^{-2}$. Let

$$\begin{aligned} L_j &= \sum_{k=1}^{n} I_{kj}, \qquad L_j^* = \sum_{k=1}^{n} \mathsf{E}(I_{kj}|\xi_{k-1}), \\ H_j &= \sum_{k=1}^{n} \mathsf{E}(I_{kj}|\xi_{k-m,k}) \quad \text{and} \quad H_j^* = \sum_{k=1}^{n} \mathsf{E}(I_{kj}|\xi_{k-m,k-1}), \end{aligned} \tag{4.7}$$

where $I_{kj} = I\{|b^{-1}X_k - t_j \pm 1| \le Cn^{-2}\}$. Then

$$\sup_{t_{j-1}\le t\le t_j} \left| \sum_{k=1}^{n} (Z_{k,t} - Z_{k,t_j}) \right| \le \frac{C}{n} + CL_j + CL_j^* + CH_j + CH_j^*. \tag{4.8}$$

Since $f_{X_n|\xi_{n-1}}(x)$ is bounded, $\mathsf{E}(I_{kj}|\xi_{k-1}) \le Cn^{-2}b$. Hence, $L_j^* \le Cn^{-1}b$ and $D_{kj} = I_{kj} - \mathsf{E}(I_{kj}|\xi_{k-1})$ satisfies $\mathsf{E}(D_{kj}^2|\xi_{k-1}) \le Cn^{-2}b$. Let $L_\diamond = \max_{1\le j\le q_n} L_j$.

Applying the inequality due to Freedman (1975) to $L_j - L_j^* = \sum_{k=1}^n D_{kj}$, we have

$$\begin{aligned} \mathsf{P}(L_\diamond \geq 9\log n) &\leq \mathsf{P}\Big(\max_{1\leq j\leq q_n} |L_j - L_j^*| \geq 8\log n\Big) + \mathsf{P}\Big(\max_{1\leq j\leq q_n} L_j^* \geq \log n\Big) \\ &\leq 2q_n \exp\left[\frac{(8\log n)^2}{-2\times(8\log n) - 2Cn^{-1}b}\right] = o(n^{-2}). \end{aligned} \tag{4.9}$$

Similarly, we have $H_j^* \leq Cn^{-1}b$, and, for $H_\diamond = \max_{1\leq j\leq q_n} H_j$, $\mathsf{P}(H_\diamond \geq 9\log n) = o(n^{-2})$. Since $\log n = o(\sqrt{nb}/(\log b^{-1})^2)$, by (4.8) and (4.9), it remains to show that

$$\mathsf{P}\left(\max_{1\leq j\leq q_n}\left|\sum_{k=1}^n Z_{k,t_j}\right| \geq 2^{-1}\sqrt{nb}(\log b^{-1})^{-2}\right) = o(1). \tag{4.10}$$

We first consider the case of X_n in (2.2). Recall (2.1) for $\xi_{k,n}^\star$. Define

$$K_{x,t}(\xi_{k-1}) = K\left(\frac{x + g(\xi_{k-1})}{b} - t\right) \quad \text{and} \quad K_{x,t}^\Delta = K_{x,t}(\xi_{k-1}) - K_{x,t}(\xi_{k-1,m}^\star).$$

Let $W_k = |g(\xi_{k-1}) - g(\xi_{k-1,m}^\star)|$. By condition (C2), $\|W_k\|_{p'} = O(m^{-\gamma})$. By Lemma 4.8, we have $\int_{-\infty}^\infty (K_{x,t}^\Delta)^2\,dx \leq Cb\min((W_k/b)^\alpha, 1)$. Hence, by Jensen's inequality,

$$\begin{aligned} \mathsf{E}(Z_{k,t}^2|\xi_{k-1}) &\leq \int_{-\infty}^\infty (K_{x,t}(\xi_{k-1}) - \mathsf{E}[K_{x,t}(\xi_{k-1})|\xi_{k-m,k-1}])^2 f_\varepsilon(x)\,dx \\ &\leq \mathsf{E}\left[\int_{-\infty}^\infty (K_{x,t}^\Delta)^2 f_\varepsilon(x)\,dx \middle| \xi_{k-m,k-1}\right] \\ &\leq Cb\mathsf{E}[\min((W_k/b)^\alpha, 1)|\xi_{k-m,k-1}]. \end{aligned} \tag{4.11}$$

Let $V = \max_{1\leq j\leq q_n}\sum_{k=1}^n \mathsf{E}(Z_{k,t_j}^2|\xi_{k-1})$. Since $\delta_1/\gamma < \tau < 1-\delta_1$ and $m \sim n^\tau$,

$$\begin{aligned} \mathsf{P}\left(V \geq \frac{nb}{(\log b^{-1})^6}\right) &\leq C(\log b^{-1})^6 \mathsf{E}\min((W_k/b)^\alpha, 1) \\ &\leq C(\log n)^6\left(\frac{\Psi_{m,p'}}{b}\right)^{\min(p',\alpha)} = o(1). \end{aligned} \tag{4.12}$$

By Freedman's (1975) inequality for martingale differences, we have

$$\begin{aligned} &\mathsf{P}\left(\max_{1\leq j\leq q_n}\left|\sum_{k=1}^n Z_{k,t_j}\right| \geq \frac{\sqrt{nb}}{2(\log b^{-1})^2}, V \leq \frac{nb}{(\log b^{-1})^6}\right) \\ &\quad \leq 2q_n \exp\left[-\frac{nb(\log b^{-1})^{-4}}{C\sqrt{nb}(\log b^{-1})^{-2} + Cnb(\log b^{-1})^{-6}}\right] = o(1) \end{aligned}$$

by condition (C1). So (4.10) follows from (4.12).

The proof of (4.10) for X_n in Theorem 2.2 is simpler. Let $p_1 = \min(p,1)$ and $\rho_1 \in (\rho,1)$. We have, by (C2)$'$ and (C3)$'$, that

$$\sup_{t\in\mathsf{R}} \mathsf{E}|Z_{k,t}| \le C\mathsf{P}(|X_k - X^\star_{k,m}| \ge \rho_1^m) + Cb^{-1}\rho_1^m$$
$$+ C\sup_{t\in\mathsf{R}} \mathsf{P}(|X_k - tb \pm b| \le \rho_1^m) \le C(\rho/\rho_1^{p_1})^m + Cb^{-1}\rho_1^m.$$

Hence, using Markov's inequality, (4.10) follows. $\square$

Proof of Lemma 4.3. Let $A = (\log b^{-1})^{-3} = o((\log b^{-1})^{-2})$. Recall the proof of Lemma 4.2 for t_j. From the proof of Lemma 4.2, we only need to consider the behavior of $R_n(t)$ at grids t_j. Note that $\tau < \tau_1$ and

$$\sup_{t\in\mathsf{R}} \sum_{j=1}^{\iota_n+1} \sum_{k\in I_j} \mathsf{E}[K^2((X_k - t)/b)|\xi_{k-1}] \le C(n^{1-\tau_1+\tau} + n^{\tau_1})b \qquad \text{a.s.} \tag{4.13}$$

By Freedman's inequality for martingale differences and (4.13),

$$\mathsf{P}\Big(\max_{0\le j\le q_n} |R_n(t_j)| \ge A\Big) \le 4q_n \exp\left[\frac{A^2 nb}{-2CA\sqrt{nb} - 2C(n^{1-\tau_1+\tau} + n^{\tau_1})b}\right] = o(1)$$

since $n^{-\delta_1} = O(b)$. Hence, (4.3) follows. $\square$

Proof of Lemma 4.4. From the proof of Lemma 4.2, we only need to show that

$$\sup_{0\le j\le q_n} |M_n(t_j)| = O_{\mathsf{P}}(\sqrt{\log n}),$$

which follows from $\sup_{t\in\mathsf{R}} \mathsf{E}[K^2((X_k - t)/b)|\xi_{k-1}] \le Cb$ a.s. and Freedman's inequality for martingale differences. $\square$

4.2. *Proof of Lemma 4.5.* As in Bickel and Rosenblatt (1973), we split the interval $[0, b^{-1}]$ into alternating big and small intervals $W_1, V_1, \dots, W_N, V_N$, where $W_i = [a_i, a_i + w]$, $V_i = [a_i + w, a_{i+1}]$, $a_i = (i-1)(w+v)$, $a_{N+1} = b^{-1}$ and $N = \lfloor b^{-1}/(w+v)\rfloor$. We will let v be sufficiently small and w be fixed. We shall first approximate $\Omega^+ := \sup_{0\le t\le b^{-1}} \widetilde{M}_n(t)$ by $\Psi^+ := \max_{1\le k\le N} \Upsilon_k^+$, where $\Upsilon_k^+ := \sup_{t\in W_k} \widetilde{M}_n(t)$, and then approximate Υ_k^+ via discretization by

$$\Xi_k^+ := \max_{1\le j\le\chi} \widetilde{M}_n(a_k + jax^{-2/\alpha}) \qquad \text{where } \chi = \lfloor wx^{2/\alpha}/a\rfloor, a > 0. \tag{4.14}$$

We similarly define Ω^-, Ψ^-, Υ_k^- and Ξ_k^- by replacing "sup" or "max" by "inf" or "min," respectively. Let $\Omega = \sup_{0\le t\le b^{-1}} |\widetilde{M}_n(t)| = \max(\Omega^+, -\Omega^-)$.

Define

$$R_1 = \mathsf{P}\Big(\max_{1\le k\le N}\sup_{t\in V_k}\widetilde{M}_n(t)\ge x\Big); \qquad R_2 = \mathsf{P}\Big(\min_{1\le k\le N}\inf_{t\in V_k}\widetilde{M}_n(t)\le -x\Big);$$

$$R_3 = \sum_{k=1}^{N}|\mathsf{P}(\Upsilon_k^+\ge x) - \mathsf{P}(\Xi_k^+\ge x)|;$$

$$R_4 = \sum_{k=1}^{N}|\mathsf{P}(\Upsilon_k^-\le -x) - \mathsf{P}(\Xi_k^-\le -x)|,$$

where $x = x_z = d_n + z/(2\log b^{-1})^{1/2}$. To deal with $R_1,\dots,R_4$, we need the following Lemma 4.6 which will be proved in Section 4.3.

Let $(\alpha, C_0) = (1, K_1)$ if $K_1 > 0$ and $(\alpha, C_0) = (2, K_2)$ if $K_1 = 0$. Let $H_\alpha(a)$ and H_α be the Pickands constants [see Theorem A1 and Lemmas A1 and A3 in Bickel and Rosenblatt (1973)]. Note that $H_1 = 1$ and $H_2 = 1/\sqrt{\pi}$.

LEMMA 4.6. *Let $t > 0$ be such that $\inf\{s^{-\alpha}(1-r(s)) : 0\le s\le t\} > 0$, where $r(s)$ is defined in Lemma 4.8. Let $\psi(x) = e^{-x^2/2}/(x\sqrt{2\pi})$. Under conditions of Theorems 2.1 or 2.2, we have for $a > 0$,*

$$\mathsf{P}\Bigg(\bigcup_{j=1}^{\lfloor tx^{2/\alpha}/a\rfloor}\{\widetilde{M}_n(v + jax^{-2/\alpha})\ge x\}\Bigg) = x^{2/\alpha}\psi(x)\frac{H_\alpha(a)}{a}C_0^{1/\alpha}t + o(x^{2/\alpha}\psi(x)) \tag{4.15}$$

uniformly over $0\le v\le b^{-1}$. The limit version of (4.15) with $a\to 0$ also holds:

$$\mathsf{P}\Bigg(\bigcup_{0\le s\le t}\{\widetilde{M}_n(v+s)\ge x\}\Bigg) = x^{2/\alpha}\psi(x)H_\alpha C_0^{1/\alpha}t + o(x^{2/\alpha}\psi(x)). \tag{4.16}$$

The left tail version of (4.15) and (4.16) also hold with "$\ge x$" replaced by "$\le -x$."

By Lemma 4.6, elementary calculations show that, for $x = x_z$,

$$\operatorname{LIM} R_j := \lim_{a\to 0}\limsup_{v\to 0}\limsup_{n\to\infty} R_j = 0, \qquad j = 1,\dots,4. \tag{4.17}$$

Note that $\Omega^+ = \max_{1\le k\le N}\sup_{t\in W_k\cup V_k}\widetilde{M}_n(t)$. By a similar identity for Ω^-, we have

$$|\mathsf{P}(\Omega\ge x) - \mathsf{P}(\{\Psi^+\ge x\}\cup\{\Psi^-\le -x\})| \le R_1 + R_2,$$

which implies $\mathrm{LIM}|\mathsf{P}(\Omega \geq x) - h(x)| = 0$ for

$$h(x) = \mathsf{P}\left(\bigcup_{k=1}^{N}\{\Xi_k^+ \geq x\} \cup \bigcup_{k=1}^{N}\{\Xi_k^- \leq -x\}\right) \tag{4.18}$$

in view of $|\mathsf{P}(\{\Psi^+ \geq x\} \cup \{\Psi^- \leq -x\}) - h(x)| \leq R_3 + R_4$. So (4.4) follows from Lemma 4.7 below which will be proved in Section 4.4.

LEMMA 4.7. *Recall (4.17) for the definition of the triple limit* LIM. *Under conditions of Theorems 2.1 or 2.2, we have* $\mathrm{LIM}|h(x_z) - (1 - e^{-2e^{-z}})| = 0$ *for all* $z \in \mathsf{R}$.

4.3. *Proof of Lemma 4.6.* We need the following lemma.

LEMMA 4.8 [Theorems B1 and B2 in Bickel and Rosenblatt (1973)]. *Under condition* (C4), *for* $r(s) = \int K(x)K(x+s)\,dx/\lambda_K$, *we have as* $s \to 0$ *that*

$$r(s) = 1 - \frac{\int (K(x) - K(x+s))^2\,dx}{2\lambda_K} = 1 - C_0|s|^\alpha + o(|s|^\alpha).$$

Now we prove Lemma 4.6. Assume $C_0 = 1$. The general case follows from a simple scale transform. Let $s_j = j/(\log n)^6$, $1 \leq j < t_n$, where $t_n = 1 + \lfloor (\log n)^6 t \rfloor$, $s_{t_n} = t$. Write $[s_{j-1}, s_j] = \bigcup_{k=1}^{q_n}[s_{j,k-1}, s_{j,k}]$, where $q_n = \lfloor (s_j - s_{j-1})n^2 \rfloor = \lfloor n^2/(\log n)^6 \rfloor$ and $s_{j,k} - s_{j,k-1} = (s_j - s_{j-1})/q_n$. Define $\Gamma_j(s) = \widetilde{M}_n(v+s) - \widetilde{M}_n(v+s_{j-1})$. Using the arguments in (4.8) and (4.9), we have

$$A_3 := \mathsf{P}\left(\max_{1 \leq k \leq q_n} \sup_{s_{j,k-1} \leq s \leq s_{j,k}} |\Gamma_j(s) - \Gamma_j(s_{j,k-1})| > \frac{(\log n)^{-2}}{2}\right) \leq \frac{C}{e^{(\log n)^2}}.$$

Let $M = 2\sqrt{nb}(\log n)^{-4}$. By truncation and Bernstein's inequality,

$$\begin{aligned}
A_2 &:= q_n \max_k \mathsf{P}(|\Gamma_j(s_{j,k})| > (\log n)^{-2}/2) \\
&\leq q_n \max_k \left[\exp\left(-\frac{Cnb(\log n)^{-4}}{B_n}\right) + \exp\left(-\frac{C\sqrt{nb}(\log n)^{-2}}{M}\right)\right] \\
&\quad + q_n \mathsf{P}\left(\left|\sum_{l=1}^{\iota_n}(u_l^\triangle - \mathsf{E}u_l^\triangle)\right| \geq \sqrt{nb}(\log n)^{-2}/4\right),
\end{aligned}$$

where $u_l^\triangle = T_l I\{|T_l| \geq \sqrt{nb}(\log n)^{-4}\}$, $T_l = u_l(v+s_{j,k}) - u_l(v+s_{j-1})$, and

$$\begin{aligned}
B_n &\leq \sum_{j=1}^{\iota_n} |H_j| \mathsf{E}(K(X_1/b - v - s_{j,k}) - K(X_1/b - v - s_{j-1}))^2 \\
&\leq \sum_{j=1}^{\iota_n} |H_j| Cb|s_{j,k} - s_{j-1}|^\alpha \leq Cnb(\log n)^{-6}.
\end{aligned}$$

Here we applied Lemma 4.8. Since $\tau_1 < 1 - \delta_1$ and $n^{-\delta_1} = O(b)$, for any $Q > 2$,

$$\mathsf{E}|u_l^{\triangle}|^2 \le C(nb)^{-Q/2}(\log n)^{4Q} n^{\tau_1(Q+2)/2} b \le Cn^{-\tau_Q}, \tag{4.19}$$

where $\tau_Q \to \infty$ as $Q \to \infty$. So $A_2 \le Cn^{-2Q}$ for any $Q > 0$, and

$$A_1 := \mathsf{P}\Big(\max_{1\le j\le t_n} \sup_{s_{j-1}<s\le s_j} |\Gamma_j(s)| > (\log n)^{-2}\Big) = \frac{O(t_n)}{n^{2Q}} \le Cn^{-Q}$$

for any $Q > 0$. Then we have the discretization approximation

$$\mathsf{P}\Big(\sup_{0\le s\le t} \widetilde{M}_n(v+s) \ge x\Big) \le \mathsf{P}\Big(\max_{1\le j\le t_n} \widetilde{M}_n(v+s_j) \ge x - (\log n)^{-2}\Big) + A_1.$$

We now apply the multivariate Gaussian approximation result in Zaĭtsev (1987) to handle $\widetilde{M}_n(v)$. To this end, we introduce

$$\begin{aligned} \widehat{M}_n(t) &= \frac{1}{\sqrt{nb\lambda_K f(bt)}} \sum_{j=1}^{\iota_n} \hat{u}_j(t) \\ &\text{where } \hat{u}_j(t) = u_j^{\diamond}(t) - \mathsf{E}u_j^{\diamond}(t), \\ &u_j^{\diamond}(t) = u_j(t) I\{|u_j(t)| \le \sqrt{nb}(\log n)^{-20}\}. \end{aligned} \tag{4.20}$$

As in (4.19), we have for any large Q,

$$\sup_t \max_{1\le j\le \iota_n} \|\hat{u}_j(t) - u_j(t)\| \le Cn^{-Q}. \tag{4.21}$$

By (4.21) and Theorem 1.1 in Zaĭtsev (1987), we have for all large Q,

$$\begin{aligned} &\mathsf{P}\Big(\max_{1\le j\le t_n} \widetilde{M}_n(v+s_j) \ge x - (\log n)^{-2}\Big) \\ &\quad\le \mathsf{P}\Big(\max_{1\le j\le t_n} \widehat{M}_n(v+s_j) \ge x - (\log n)^{-2}\Big) + Cn^{-Q} \\ &\quad\le \mathsf{P}\Big(\max_{1\le j\le t_n} Y_n(j) \ge x_n'\Big) + Ct_n^{5/2} \exp\bigg(-\frac{C(\log n)^{18}}{t_n^{5/2}}\bigg) + Cn^{-Q}, \end{aligned} \tag{4.22}$$

where $x_n' = x - 2(\log n)^{-2}$ and $(Y_n(1), \ldots, Y_n(t_n))$ is a centered Gaussian random vector with covariance matrix

$$\widehat{\Sigma}_n = \mathsf{Cov}(\widehat{M}_n(v+s_1), \ldots, \widehat{M}_n(v+s_{t_n})). \tag{4.23}$$

By Lemma 4.9 below and Lemma A4 in Bickel and Rosenblatt (1973), we have

$$\begin{aligned} \mathsf{P}\Big(\max_{1\le j\le t_n} Y_n(j) \ge x_n'\Big) &\le \mathsf{P}\Big(\max_{1\le j\le t_n} \widetilde{Y}_n(s_j) \ge x_n'\Big) + \frac{Ct_n^2(t_n^2(b+n^{-\varpi}))^{1/2}}{\exp({x_n'}^2/2)} \\ &\le \mathsf{P}\Big(\max_{1\le j\le t_n} \widetilde{Y}_n(s_j) \ge x_n'\Big) + Cb^{1+\delta} \end{aligned}$$

for some $\delta > 0$, where $\widetilde{Y}_n(\cdot)$ is a separable stationary Gaussian process with mean 0 and covariance function $r(\cdot)$. By Lemma A3 in Bickel and Rosenblatt (1973) and some elementary calculations,

$$\begin{aligned}\mathsf{P}\Big(\max_{1\le j\le t_n}\widetilde{Y}_n(s_j)\ge x_n'\Big)&\le \mathsf{P}\Big(\sup_{0\le s\le t}\widetilde{Y}_n(s)\ge x_n'\Big)\\&=x^{2/\alpha}\psi(x)H_\alpha t+o(x^{2/\alpha}\psi(x)).\end{aligned}$$

This implies the upper bound in (4.16). With the same argument, for any $a > 0$,

$$\begin{aligned}&\mathsf{P}\Big(\sup_{0\le s\le t}\widetilde{M}_n(v+s)\ge x\Big)\\&\quad\ge \mathsf{P}\Bigg(\bigcup_{j=1}^{[tx^{2/\alpha}/a]}\{\widetilde{M}_n(v+jax^{-2/\alpha})\ge x\}\Bigg)\\&\quad\ge \mathsf{P}\Bigg(\bigcup_{j=1}^{[tx^{2/\alpha}/a]}\widetilde{Y}_n(jax^{-2/\alpha})\ge x+2(\log n)^{-2}\Bigg)-Cb^{1+\delta}\\&\quad\ge \mathsf{P}\Bigg(\bigcup_{j=1}^{[tx^{2/\alpha}/a]}\widetilde{Y}_n(jax^{-2/\alpha})\ge x\Bigg)\\&\qquad-\sum_{j=1}^{[tx^{2/\alpha}/a]}\mathsf{P}(x\le\widetilde{Y}_n(jax^{-2/\alpha})<x+2(\log n)^{-2})-Cb^{1+\delta}\\&\quad=x^{2/\alpha}\psi(x)\frac{H_\alpha(a)}{a}t+o(x^{2/\alpha}\psi(x)).\end{aligned}$$

Then the low bound in (4.16) is obtained by (A20) in Bickel and Rosenblatt (1973), letting first $n\to\infty$ and then $a\to 0$.

Using a similar and simpler proof, we can prove (4.15).

LEMMA 4.9. *For the covariance matrix $\widehat{\Sigma}_n$ defined in (4.23), we have*

$$|\widehat{\Sigma}_n-(r(s_j-s_i))_{1\le i,j\le t_n}|\le Ct_n^2(b+n^{-\varpi})\qquad\textit{for some } \varpi>0. \tag{4.24}$$

PROOF. Let $\Sigma_n = \mathsf{Cov}(\widetilde{M}_n(v+s_1),\dots,\widetilde{M}_n(v+s_{t_n}))$. By (4.21), $|\Sigma_n - \widehat{\Sigma}_n|\le Cn^{-Q}$ for any $Q>0$. Note that $\mathsf{E}(R_n^2(t))\le Cn^{\tau-\tau_1}$ and $\tau_1>\tau$. Then

$$|\mathsf{Cov}(\widetilde{M}_n(s),\widetilde{M}_n(t))-\mathsf{Cov}(\widetilde{M}_n(s)+R_n(s),\widetilde{M}_n(t)+R_n(t))|\le Cn^{\tau/2-\tau_1/2}.$$

By (4.11), we obtain that $\|\widetilde{M}_n(t)+R_n(t)-M_n(t)\|^2\le Cn^{\delta_1-\tau\gamma}$. Thus,

$$|\mathsf{Cov}(M_n(s),M_n(t))-\mathsf{Cov}(\widetilde{M}_n(s)+R_n(s),\widetilde{M}_n(t)+R_n(t))|\le Cn^{\delta_1/2-\tau\gamma/2}.$$

Since $K(x)=0$ if $|x|>1$, for $0\le s,t\le b^{-1}$, we have

$$|\mathsf{E}[K(X_k/b-s)K(X_k/b-t)]-b\sqrt{f(bs)f(bt)}r(s-t)\lambda_K|\le Cb^2.$$

Note that $\mathsf{E}(|K(X_k/b-t)||\xi_{k-1})\le Cb$. Therefore,

$$|\mathsf{Cov}(M_n(s),M_n(t))-r(s-t)|\le Cb.$$

Combining the above arguments, we prove (4.24). $\square$

4.4. *Proof of Lemma 4.7.* Let $\widehat{M}_n(t)$ be defined in (4.20) with 20 therein replaced by $20d$. Also, d may vary accordingly. Let $x_n=x\pm(\log n)^{-2d}$ and

$$\begin{aligned}
\mathbf{B}_{k,j}&=\{\widetilde{M}_n(a_k+jax^{-2/\alpha})\ge x\}\cup\{\widetilde{M}_n(a_k+jax^{-2/\alpha})\le -x\},\\
\widehat{\mathbf{B}}^{\pm}_{k,j}&=\{\widehat{M}_n(a_k+jax^{-2/\alpha})\ge x_n\}\cup\{\widehat{M}_n(a_k+jax^{-2/\alpha})\le -x_n\},\\
\mathbf{D}_{k,j}&=\{Y_n(a_k+jax^{-2/\alpha})\ge x\}\cup\{Y_n(a_k+jax^{-2/\alpha})\le -x\},\\
\mathbf{D}^{\pm}_{k,j}&=\{Y_n(a_k+jax^{-2/\alpha})\ge x_n\}\cup\{Y_n(a_k+jax^{-2/\alpha})\le -x_n\},\\
\widehat{\mathbf{D}}^{\pm}_{k,j}&=\{\widehat{Y}_n(a_k+jax^{-2/\alpha})\ge x_n\}\cup\{\widehat{Y}_n(a_k+jax^{-2/\alpha})\le -x_n\},
\end{aligned}$$

where $Y_n(\cdot)$ and $\widehat{Y}_n(\cdot)$ are centered Gaussian processes with covariance functions

$$\begin{aligned}
\mathsf{Cov}(Y_n(s_1),Y_n(s_2))&=\mathsf{Cov}(\widetilde{M}_n(s_1),\widetilde{M}_n(s_2)),\\
\mathsf{Cov}(\widehat{Y}_n(s_1),\widehat{Y}_n(s_2))&=\mathsf{Cov}(\widehat{M}_n(s_1),\widehat{M}_n(s_2)),
\end{aligned}$$

respectively. Recall (4.14) for χ. Let

$$\mathbf{A}_k=\bigcup_{j=1}^{\chi}\mathbf{B}_{k,j},\qquad \mathbf{C}_k=\bigcup_{j=1}^{\chi}\mathbf{D}_{k,j},\qquad \mathbf{C}^{\pm}_k=\bigcup_{j=1}^{\chi}\mathbf{D}^{\pm}_{k,j}\quad\text{and}\quad \widehat{\mathbf{C}}^{\pm}_k=\bigcup_{j=1}^{\chi}\widehat{\mathbf{D}}^{\pm}_{k,j}.$$

LEMMA 4.10. *Let $N=\lfloor b^{-1}/(w+v)\rfloor$. Under the conditions of Theorems 2.1 or 2.2, we have for any fixed integer l satisfying $1\le l\le N/2$ that*

$$\left|\mathsf{P}\left(\bigcup_{k=1}^{N}\mathbf{A}_k\right)-\sum_{d=1}^{2l-1}(-1)^{d-1}\left(\sum_{1\le i_1<\cdots<i_d\le N}-\sum_{\mathcal{I}}\right)\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{C}_{i_j}\right)\right|\le\frac{C_1^{2l}}{(2l)!}+\frac{O(1)}{\log n},$$

where C_1 does not depend on l, and $\mathcal{I}$ is defined in (4.26).

PROOF. By Bonferroni's inequality, we have

$$\sum_{d=1}^{2l}(-1)^{d-1}\sum_{1\le i_1<\cdots<i_d\le N}\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{A}_{i_j}\right)$$

$$\leq \mathsf{P}\left(\bigcup_{k=1}^{N}\mathbf{A}_k\right) \leq \sum_{d=1}^{2l-1}(-1)^{d-1}\sum_{1\leq i_1<\cdots<i_d\leq N}\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{A}_{i_j}\right). \tag{4.25}$$

We now estimate the probability $\mathsf{P}(\bigcap_{j=1}^{d}\mathbf{A}_{i_j})$. Recall $W_k = [a_k, a_k + w)$. Let $q_j = i_{j+1} - i_j$, $1 \leq j \leq d-1$. Define the index set

$$\mathcal{I} := \left\{1 \leq i_1 < \cdots < i_d \leq N : \min_{1\leq j\leq d-1} q_j \leq \lfloor 2w^{-1} + 2\rfloor\right\}. \tag{4.26}$$

Let $0 \leq d_0 \leq d-2$ and

$$\mathcal{I}_{d_0} = \{1 \leq i_1 < \cdots < i_d \leq N : \text{the number of } j \text{ such that } q_j > \lfloor 2w^{-1} + 2\rfloor \text{ is } d_0\}.$$

Then we have $\mathcal{I} = \bigcup_{d_0=0}^{d-2}\mathcal{I}_{d_0}$. We can see that the number of elements in the sum $\sum_{\mathcal{I}_{d_0}}\mathsf{P}(\bigcap_{j=1}^{d}\mathbf{A}_{i_j})$ is bounded by $CN^{d_0+1} = O(b^{-d_0-1})$, where C is independent of N. Suppose now $i_1,\dots,i_d$ are in $\mathcal{I}_{d_0}$. Write

$$\bigcap_{j=1}^{d}\mathbf{A}_{i_j} = \bigcup_{j_1=1}^{\chi}\cdots\bigcup_{j_d=1}^{\chi}\{\mathbf{B}_{i_1,j_1}\cap\cdots\cap\mathbf{B}_{i_d,j_d}\}.$$

Without loss of generality, we assume $q_1 \leq \lfloor 2w^{-1}+2\rfloor$, $q_2 > \lfloor 2w^{-1}+2\rfloor,\dots,$ $q_{d_0+1} > \lfloor 2w^{-1}+2\rfloor$. By (4.21) and Theorem 1.1 in Zaĭtsev (1987), we have for all large Q,

$$\begin{aligned}\mathsf{P}(\mathbf{B}_{i_1,j_1}\cap\cdots\cap\mathbf{B}_{i_d,j_d}) &\leq \mathsf{P}(\widehat{\mathbf{B}}^{-}_{i_1,j_1}\cap\cdots\cap\widehat{\mathbf{B}}^{-}_{i_d,j_d}) + Cn^{-Q}\\ &\leq \mathsf{P}(\widehat{\mathbf{D}}^{-}_{i_1,j_1}\cap\cdots\cap\widehat{\mathbf{D}}^{-}_{i_d,j_d}) + C\exp(-(\log b^{-1})^2) + Cn^{-Q}.\end{aligned} \tag{4.27}$$

By (4.21), we have uniformly in s_1 and s_2 that, for any large Q,

$$|\mathsf{Cov}(Y_n(s_1), Y_n(s_2)) - \mathsf{Cov}(\widehat{Y}_n(s_1), \widehat{Y}_n(s_2))| \leq Cn^{-Q}. \tag{4.28}$$

Using the argument of (4.24), there exists $C > 0$ and $\varpi > 0$, such that for $\nu_n = C(b + n^{-\varpi})$ and any $1 \leq j_{(\cdot)} \leq \chi$, we have

$$\begin{aligned}&|\mathsf{Cov}(Y_n(a_{i_l} + j_l a x^{-2/\alpha}), Y_n(a_{i_k} + j_k a x^{-2/\alpha}))| \leq \nu_n\\ &\qquad\qquad\qquad \text{for } 3 \leq k \leq d_0+1, l = 1,2;\\ &|\mathsf{Cov}(Y_n(a_{i_s} + j_s a x^{-2/\alpha}), Y_n(a_{i_k} + j_k a x^{-2/\alpha}))| \leq \nu_n \qquad \text{for } 3 \leq k \neq s \leq d_0+1;\\ &\qquad |\mathsf{Var}(Y_n(a_{i_k} + j_k a x^{-2/\alpha})) - 1| \leq \nu_n \qquad \text{for } 1 \leq k \leq d_0+1;\end{aligned}$$

and, letting $\mu = r(a_{i_2} - a_{i_1} + (j_2 - j_1)a x^{-2/\alpha})$,

$$|\mathsf{Cov}(Y_n(a_{i_1} + j_1 a x^{-2/\alpha}), Y_n(a_{i_2} + j_2 a x^{-2/\alpha})) - \mu| \leq \nu_n.$$

Note that $|j_2 - j_1| a x^{-2/\alpha} \le w$ and $a_{i_2} - a_{i_1} \ge w + v$ and $\sup_{x \ge v} |r(x)| < 1$. Let any $1 \le j_{(\cdot)} \le \chi$ and $\mathbf{V}_n$ be the covariance matrix of the Gaussian vector $(\widehat{Y}_1, \ldots, \widehat{Y}_{d_0+1})$, where $\widehat{Y}_k = \widehat{Y}_n(a_{i_k} + j_k a x^{-2/\alpha})$, $1 \le k \le d$. Using the bounds of the covariances above, we have for some $\delta > 0$ that

$$|\mathbf{V}_n - \mathbf{V}| \le C n^{-\delta} \qquad \text{where } \mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & 0 \\ 0 & \mathbf{I}_{d_0-1} \end{pmatrix} \text{ and } \mathbf{V}_1 = \begin{pmatrix} 1 & \mu \\ \mu & 1 \end{pmatrix}. \tag{4.29}$$

By (4.29), we have

$$|\mathbf{V}_n^{-1} - \mathbf{V}^{-1}| \le C n^{-\delta} \quad \text{and} \quad |\sqrt{\det(\mathbf{V})} - \sqrt{\det(\mathbf{V}_n)}| \le C n^{-\delta}. \tag{4.30}$$

Let $p_n(y)$ be the density of $(\widehat{Y}_1, \ldots, \widehat{Y}_{d_0+1})$, and $p(y)$ be the density of the Gaussian random vector with covariance matrix $\mathbf{V}$. By (4.30), we have

$$\begin{aligned} |p_n(y) - p(y)| &\le C n^{-\delta} p(y) + C \exp(-y \mathbf{V}^{-1} y'/2) |\exp(C n^{-\delta} |y|^2) - 1| \\ &\le C(n^{-\delta} + n^{-\delta} (\log n)^2) p(y) + C \exp(-(\log n)^2 / C). \end{aligned} \tag{4.31}$$

Hereafter, $\delta > 0$ may be different in different places. Note that

$$|\mu| \le \sup_{x \ge v} |r(x)| < 1.$$

Then it follows from Lemma 2 in Berman (1962) that, for some $\delta > 0$, we have

$$\begin{aligned} &\mathsf{P}(\widehat{\mathbf{D}}^-_{i_1,j_1} \cap \cdots \cap \widehat{\mathbf{D}}^-_{i_d,j_d}) \\ &\quad \le (1 + C n^{-\delta}) \int_{\Xi^-} p(y)\, dy + C \exp(-(\log n)^2 / C) \\ &\quad \le C b^{d_0+1+\delta}, \end{aligned} \tag{4.32}$$

where $y = (y_1, \ldots, y_{d_0+1})$ and

$$\Xi^{\pm} = \bigcap_{j=1}^{d_0+1} [\{y_j \ge x_n\} \cup \{y_j \le -x_n\}].$$

Noting that $\chi^d = O(b^{-\delta/2})$ and by (4.27) and (4.32), we have for some $\delta > 0$,

$$\sum_{d_0=0}^{d-2} \sum_{\mathcal{I}_{d_0}} \mathsf{P}\left(\bigcap_{j=1}^{d} \mathbf{A}_{i_j} \right) \le C b^{\delta}. \tag{4.33}$$

We now estimate

$$\left(\sum_{1 \le i_1 < \cdots < i_d \le N} - \sum_{\mathcal{I}} \right) \mathsf{P}\left(\bigcap_{j=1}^{d} \mathbf{A}_{i_j} \right). \tag{4.34}$$

Suppose that $i_1, \dots, i_d \notin \mathcal{I}$. Since $i_{j+1} - i_j > \lfloor 2/w + 2 \rfloor$, we have $a_{i_{j+1}} - a_{i_j} \geq (w+v)\lfloor 2/w + 2 \rfloor > 2 + w + v$. Then, for $1 \leq s \neq k \leq d$, $1 \leq j_s, j_k \leq \chi$,

$$|\mathsf{Cov}(Y_n(a_{i_s} + j_s a x^{-2/\alpha}), Y_n(a_{i_k} + j_k a x^{-2/\alpha}))| \leq C(b + n^{-\varpi})$$

holds for some $\varpi > 0$. By the bounds of the covariances above, the covariance matrix $\widetilde{\mathbf{V}}_n$ of $(\widehat{Y}_1, \dots, \widehat{Y}_d)$ when $i_1, \dots, i_d \notin \mathcal{I}$ satisfies

$$|\widetilde{\mathbf{V}}_n - \mathbf{I}| \leq C n^{-\delta} \qquad \text{for some } \delta > 0. \tag{4.35}$$

For the probability in the sum in (4.34), as in (4.27) and (4.32), we have for n large,

$$\begin{aligned}
\mathsf{P}\left(\bigcap_{j=1}^{d} \mathbf{A}_{i_j}\right) &\leq \sum_{j_1=1}^{\chi} \cdots \sum_{j_d=1}^{\chi} \mathsf{P}(\mathbf{B}_{i_1,j_1} \cap \cdots \cap \mathbf{B}_{i_d,j_d}) \\
&\leq \sum_{j_1=1}^{\chi} \cdots \sum_{j_d=1}^{\chi} \mathsf{P}(\widehat{\mathbf{D}}^{-}_{i_1,j_1} \cap \cdots \cap \widehat{\mathbf{D}}^{-}_{i_d,j_d}) + C n^{-Q} \\
&\leq 2^d \sum_{j_1=1}^{\chi} \cdots \sum_{j_d=1}^{\chi} (x^{-1}\exp(-x^2/2))^d + C b^{d+\delta} + C n^{-Q} \\
&\leq 2^d (\chi x^{-1} \exp(-x^2/2))^d + C b^{1+\delta} \leq C_1^d b^d + C b^{d+\delta}
\end{aligned}$$

for some $C_1 > 0$ which does not depend on d. This together with (4.33) implies that

$$\sum_{1 \leq i_1 < \cdots < i_d \leq N} \mathsf{P}\left(\bigcap_{j=1}^{d} \mathbf{A}_{i_j}\right) \leq C_1^d / d! + C b^{\delta} \tag{4.36}$$

for some $C_1 > 0$ which does not depend on d. To prove Lemma 4.10, by (4.25), (4.33) and (4.36), we only need to show that, for $i_1, \dots, i_d \notin \mathcal{I}$,

$$\left|\mathsf{P}\left(\bigcap_{j=1}^{d} \mathbf{A}_{i_j}\right) - \mathsf{P}\left(\bigcap_{j=1}^{d} \mathbf{C}_{i_j}\right)\right| \leq C b^d (\log n)^{-d}. \tag{4.37}$$

By (4.21) and Theorem 1.1 in Zaĭtsev (1987), as in (4.22), it suffices to show

$$\left|\mathsf{P}\left(\bigcap_{j=1}^{d} \mathbf{C}_{i_j}\right) - \mathsf{P}\left(\bigcap_{j=1}^{d} \widehat{\mathbf{C}}^{\pm}_{i_j}\right)\right| \leq C b^d (\log n)^{-d}.$$

By (4.28) and Lemma A4 in Bickel and Rosenblatt (1973), using $\mathsf{P}(\bigcap_{j=1}^{d} \widehat{\mathbf{C}}^{\pm}_{i_j}) = 1 - \mathsf{P}(\bigcup_{j=1}^{d} \widehat{\mathbf{C}}^{\pm c}_{i_j})$ and the inclusion–exclusion principle, we have for any large

Q,

$$\left|\mathsf{P}\left(\bigcap_{j=1}^{d}\widehat{\mathbf{C}}_{i_j}^{\pm}\right)-\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{C}_{i_j}^{\pm}\right)\right|\le C\chi^2 n^{-2Q}\le Cn^{-Q}.$$

So it suffices to show that

$$\left|\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{C}_{i_j}\right)-\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{C}_{i_j}^{\pm}\right)\right|\le Cb^d(\log n)^{-d}. \tag{4.38}$$

By (4.35) and a similar inequality as (4.31), we have, for some $\delta>0$,

$$|\mathsf{P}(\mathbf{D}_{i_1,j_1}^{\pm}\cap\cdots\cap\mathbf{D}_{i_d,j_d}^{\pm})-(\mathsf{P}(\mathbf{D}^{\pm}))^d|\le Cb^{d+\delta},$$

where $\mathbf{D}^{\pm}=\{\mathbf{N}\ge x_n\}\cup\{\mathbf{N}\le -x_n\}$ and $\mathbf{N}$ is a standard normal random variable. It follows that, for some $\delta>0$,

$$\begin{aligned}
&\left|\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{C}_{i_j}^{-}\right)-\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{C}_{i_j}^{+}\right)\right|\\
&\quad\le\sum_{j_1=1}^{\chi}\cdots\sum_{j_d=1}^{\chi}|\mathsf{P}(\mathbf{D}_{i_1,j_1}^{-}\cap\cdots\cap\mathbf{D}_{i_d,j_d}^{-})-\mathsf{P}(\mathbf{D}_{i_1,j_1}^{+}\cap\cdots\cap\mathbf{D}_{i_d,j_d}^{+})|\\
&\quad=\sum_{j_1=1}^{\chi}\cdots\sum_{j_d=1}^{\chi}|(\mathsf{P}(\mathbf{D}^{-}))^d-(\mathsf{P}(\mathbf{D}^{+}))^d|+Cb^{d+\delta}.
\end{aligned}$$

So (4.38) follows from $\mathsf{P}(\mathbf{D}^-)-\mathsf{P}(\mathbf{D}^+)\le C(\log n)^{-2d}b$ and $\mathsf{P}(\mathbf{D}^{\pm})\le Cb/(\log b^{-1})^{1/\alpha}$. The lemma is then proved. $\square$

We are ready to prove Lemma 4.7. Let $\{\varepsilon_i^{(k)}\}_{i\in\mathsf{Z}}$, $1\le k\le n$, be i.i.d. copies of $\{\varepsilon_i\}_{i\in\mathsf{Z}}$, and $\xi_j^{(k)}=(\ldots,\varepsilon_{j-1}^{(k)},\varepsilon_j^{(k)})$. Let $X_j^{(k)}=G(\xi_j^{(k)})$. Then $X_k^{(k)}$, $1\le k\le n$, are i.i.d. Now define $\mathbf{A}_k'$, $M_n'(t)$, $\widetilde{M}_n'(t)$, $N_n'(t)$, $R_n'(t)$, $R_1',\ldots,R_4'$ by replacing X_k and $\{\varepsilon_i\}$ by $X_k^{(k)}$ and $\{\varepsilon_i^{(k)}\}$, respectively, in the above proofs. Repeating the arguments above, we can obtain that

$$\left|\mathsf{P}\left(\bigcup_{k=1}^{N}\mathbf{A}_k'\right)-\sum_{d=1}^{2l-1}(-1)^{d-1}\left(\sum_{1\le i_1<\cdots<i_d\le N}-\sum_{\mathcal{I}}\right)\mathsf{P}\left(\bigcap_{j=1}^{d}\mathbf{C}_{i_j}\right)\right|\le\frac{C_1^{2l}}{(2l)!}+\frac{O(1)}{\log n}.$$

By letting $n\to\infty$ and then $l\to\infty$, we have

$$\limsup_{n\to\infty}\left|\mathsf{P}\left(\bigcup_{k=1}^{N}\mathbf{A}_k\right)-\mathsf{P}\left(\bigcup_{k=1}^{N}\mathbf{A}_k'\right)\right|=0.$$

Similarly, (4.17) holds with R_j therein replaced by R'_j. Hence, as $n \to \infty$,

$$\text{LIM}\left|\mathsf{P}\left(\bigcup_{k=1}^{N} \mathbf{A}'_k\right) - \mathsf{P}\Big(\sup_{0\le t\le b^{-1}} |\widetilde{M}'_n(t)| < x\Big)\right| = 0. \tag{4.39}$$

Note that Lemmas 4.1–4.3 also hold for $(X_k^{(k)})_{k\in\mathsf{Z}}$, $M'_n(t)$, $\widetilde{M}'_n(t)$, $N'_n(t)$, $R'_n(t)$. By the theorem in Rosenblatt (1976), the second probability in (4.39) converges to $e^{-2e^{-z}}$. This completes the proof.

5. Proofs of Proposition 2.1, Theorems 2.4 and 2.5. Without loss of generality, we assume $l = 0$, $u = 1$. We first introduce the truncation

$$\breve{Z}_k = Z_k I\{|Z_k| \le (\log n)^{12/(p-2)}\} - \mathsf{E}(Z_k I\{|Z_k| \le (\log n)^{12/(p-2)}\}),$$
$$\widetilde{Z}_k = Z_k I\{|Z_k| > \sqrt{nb}/(\log n)^4\} - \mathsf{E}(Z_k I\{|Z_k| > \sqrt{nb}/(\log n)^4\})$$

and $\widehat{Z}_k = Z_k - \breve{Z}_k$, $1 \le k \le n$. Correspondingly, define

$$r_n(x) = \frac{1}{\sqrt{nb}} \sum_{k=1}^{n} K\left(\frac{X_k}{b} - x\right)\widehat{Z}_k =: \frac{1}{\sqrt{nb}} \sum_{k=1}^{n} w_{n,k}(x),$$
$$r_{n,1}(x) = \frac{1}{\sqrt{nb}} \sum_{k=1}^{n} K\left(\frac{X_k}{b} - x\right)\widetilde{Z}_k =: \frac{1}{\sqrt{nb}} \sum_{k=1}^{n} w_{n,k1}(x),$$
$$r_{n,2}(x) = r_n(x) - r_{n,1}(x) =: \frac{1}{\sqrt{nb}} \sum_{k=1}^{n} w_{n,k2}(x).$$

LEMMA 5.1. *Under the conditions of Proposition 2.1, we have*

$$\mathsf{P}\Big(\sup_{0\le x\le b^{-1}} |r_n(x)| \ge 3(\log n)^{-2}\Big) = o(1).$$

PROOF. Since $b \ge Cn^{-\delta_1}$ and $\mathsf{E}|Z_1|^p < \infty$, $p > 2/(1-\delta_1)$, for n large, we have

$$\begin{aligned}\mathsf{E} \sup_{0\le x\le b^{-1}} |r_{n,1}(x)| &\le Cn(nb)^{-p/2}(\log n)^{4p-4} \\ &\le Cn^{1-p(1-\delta_1)/2}(\log n)^{4p-4} \le (\log n)^{-3}.\end{aligned} \tag{5.1}$$

We now deal with $r_{n,2}$. Let $q_n = \lfloor n^2/b \rfloor$, $t_j = j/(bq_n)$, $j = 0, \dots, q_n$. As in (4.8), we have

$$\max_{0\le j\le q_n} \sup_{t_j\le t\le t_{j+1}} |r_{n,2}(t) - r_{n,2}(t_j)| \le \frac{C}{n(\log n)^4} + C\frac{\max_{0\le j\le q_n} L_j}{(\log n)^4}. \tag{5.2}$$

By (4.9), (5.1), (5.2) and since $r_{n,2}(x) + r_{n,1}(x) = r_n(x)$, it suffices to show

$$\mathsf{P}\Big(\max_{0\le j\le q_n} |r_{n,2}(t_j)| \ge 2(\log n)^{-2}\Big) = o(1). \tag{5.3}$$

Note that $\mathsf{E}(\widehat{Z}_k^2) \le C(\log n)^{-12}$. By (C3) [or (C3)$'$], we have

$$\max_{0\le j\le q_n} \sum_{k=1}^{n} \mathsf{E}[w_{n,k2}^2(t_j)|\widetilde{\xi}_{k-2}] \le Cnb(\log n)^{-6}. \tag{5.4}$$

Thus, (5.3) follows from (5.4) and applying Freedman's inequality to martingale differences $\{w_{n,k2}(x), k = 1,3,\ldots\}$ and $\{w_{n,k2}(x), k = 2,4,\ldots\}$. $\square$

Proof of Proposition 2.1. Let $m = \lfloor n^\tau \rfloor$, where $\delta_1/\gamma < \tau < 1 - \delta_1$, and

$$Z_k(t) = \breve{Z}_k\bigg\{K\bigg(\frac{X_k}{b} - t\bigg) - \mathsf{E}\bigg[K\bigg(\frac{X_k}{b} - t\bigg)\bigg|\xi_{k-m,k}\bigg]\bigg\}, \qquad 1 \le k \le n.$$

Note that $\{Z_1(t), Z_3(t),\ldots\}$ and $\{Z_2(t), Z_4(t),\ldots\}$ are two sequences of martingale differences. As in the proof of Lemma 4.2, we can show that

$$\begin{aligned}
&\mathsf{P}\Bigg(\sup_{0\le t\le b^{-1}}\Bigg|\sum_{k=1}^{n/2} Z_{2k-1}(t)\Bigg| \ge \sqrt{nb}(\log n)^{-2}\Bigg) = o(1),\\
&\mathsf{P}\Bigg(\sup_{0\le t\le b^{-1}}\Bigg|\sum_{k=1}^{n/2} Z_{2k}(t)\Bigg| \ge \sqrt{nb}(\log n)^{-2}\Bigg) = o(1).
\end{aligned} \tag{5.5}$$

Set

$$\widetilde{N}_n(t) = \frac{1}{\sqrt{nb\lambda_K f(bt)}} \sum_{k=1}^{n} \mathsf{E}\bigg[K\bigg(\frac{X_k}{b} - t\bigg)\bigg|\xi_{k-m,k-1}\bigg]\breve{Z}_k.$$

Since $\sup_t \mathsf{E}(\{\breve{Z}_k \mathsf{E}[K(X_k/b - t)|\xi_{k-m,k-1}]\}^2|\widetilde{\xi}_{k-1}) \le Cb^2$, we have by Freedman's inequality for martingale differences,

$$\mathsf{P}\Big(\max_{0\le j\le q_n} |\widetilde{N}_n(t_j)| \ge (\log n)^{-2}\Big) = o(1),$$

which, together with the discretization approximation as in (4.8), yields that

$$\mathsf{P}\Big(\sup_{0\le t\le b^{-1}} |\widetilde{N}_n(t)| \ge 2(\log n)^{-2}\Big) = o(1). \tag{5.6}$$

Set $\breve{\sigma}_n^2 = \mathsf{E}\breve{Z}_n^2$ and

$$\begin{aligned}
\widetilde{M}_n(t) &= \frac{1}{\sqrt{nb\lambda_K f(bt)}}\\
&\quad\times \sum_{k=1}^{n}\bigg\{\mathsf{E}\bigg[K\bigg(\frac{X_k}{b} - t\bigg)\bigg|\xi_{k-m,k}\bigg] - \mathsf{E}\bigg[K\bigg(\frac{X_k}{b} - t\bigg)\bigg|\xi_{k-m,k-1}\bigg]\bigg\}\frac{\breve{Z}_k}{\breve{\sigma}_n}.
\end{aligned}$$

Following the argument of Lemma 4.5 and replacing the truncation levels $(\log n)^{-20}$ and $(\log n)^{-20d}$ in (4.20) and the proof of Lemma 4.7 with $(\log n)^{-20p/(p-2)}$ and $(\log n)^{-20pd/(p-2)}$, respectively, we can get

$$\mathsf{P}\biggl((2\log b^{-1})^{1/2}\Bigl(\sup_{0\le t\le b^{-1}}|\widetilde{M}_n(t)|-d_n\Bigr)\le z\biggr)\to e^{-2e^{-z}}. \tag{5.7}$$

Note that $|1-\breve{\sigma}_n^2/\sigma^2|=O((\log n)^{-12})$. The proposition follows from Lemma 5.1 and (5.5)–(5.7). $\square$

Proof of Theorem 2.4. Write $(\mu_n(x)-\mu(x))f_n(x)=R_n^r(x)+M_{n1}^r(x)$, where

$$R_n^r(x)=\frac{1}{nb}\sum_{k=1}^n K\left(\frac{X_k-x}{b}\right)(\mu(X_k)-\mu(x)),$$

$$M_{n1}^r(x)=\frac{1}{nb}\sum_{k=1}^n K\left(\frac{X_k-x}{b}\right)\sigma(X_k)\eta_k.$$

Then Theorem 2.4 follows from Lemmas 4.4, 5.2 and 5.3 and Proposition 2.1. $\square$

Lemma 5.2. *Under the conditions of Theorem 2.4, we have*

$$\sup_{0\le x\le 1}|R_n^r(x)-b^2\psi_K\rho_\mu(x)|=O_{\mathsf{P}}(\tau_n)\qquad \textit{where } \tau_n=\sqrt{\frac{b\log n}{n}}+b^4+\frac{\mathcal{Z}_n^{1/2}b}{n}.$$

Proof. Set $\gamma_k(x)=K((X_k-x)/b)(\mu(X_k)-\mu(x))$. Let $q_n=\lfloor n^2/b\rfloor$, $t_j=j/q_n$, $j=0,\dots,q_n$. Since $\mu(\cdot)\in\mathcal{C}^4(T^\epsilon)$, $\max_{0\le j\le q_n}\mathsf{E}[\gamma_k^2(t_j)|\xi_{k-1}]\le Cb^3$. By Freedman's inequality for martingale differences, we have

$$\max_{0\le j\le q_n}\left|\sum_{k=1}^n(\gamma_k(t_j)-\mathsf{E}[\gamma_k(t_j)|\xi_{k-1}])\right|=O_{\mathsf{P}}(\sqrt{nb^3\log n}),$$

where we used the condition $0<\delta_1<1/3$. Recall that $K(x)$ and $m(x)$ are Lipschitz continuous in $[-1,1]$. Using the discretization approximation as in (4.8) and the argument in (4.9), it can be seen that

$$\sup_{0\le x\le 1}\left|\sum_{k=1}^n(\gamma_k(x)-\mathsf{E}[\gamma_k(x)|\xi_{k-1}])\right|=O_{\mathsf{P}}(\sqrt{nb^3\log n}).$$

The rest of the proof is the same as that of Lemma 2(ii) in Zhao and Wu (2008). $\square$

LEMMA 5.3. *Under the conditions of Theorem 2.4, we have*

$$\sup_{0\le x\le 1}\left|M_{n1}^r(x)-\frac{1}{nb}\sum_{k=1}^n K\left(\frac{X_k-x}{b}\right)\sigma(x)\eta_k\right| = O_{\mathsf{P}}\left(\sqrt{\frac{b\log n}{n}}\right).$$

PROOF. Let

$$\widetilde{\eta}_k=\eta_k I\{|\eta_k|\ge\sqrt{nb}/(\log n)^4\}-\mathsf{E}(\eta_k I\{|\eta_k|\ge\sqrt{nb}/(\log n)^4\}),$$

$$\widetilde{w}_{nk}(x)=K\left(\frac{X_k-x}{b}\right)(\sigma(X_k)-\sigma(x))\widetilde{\eta}_k,$$

$$\widehat{w}_{nk}(x)=K\left(\frac{X_k-x}{b}\right)(\sigma(X_k)-\sigma(x))\widehat{\eta}_k,\qquad \widehat{\eta}_k=\eta_k-\widetilde{\eta}_k.$$

Note that $\sup_{x\in T^\epsilon}|K((X_k-x)/b)(\sigma(X_k)-\sigma(x))|\le Cb$. Then

$$\mathsf{E}\sup_{x\in\mathsf{R}}\left|\frac{1}{nb}\sum_{k=1}^n\widetilde{w}_{nk}(x)\right|=O\left(\sqrt{\frac{b}{n(\log n)^4}}\right).$$

Since $\sup_{x\in\mathsf{R}}\mathsf{E}[\widehat{w}_{nk}^2(x)|\widetilde{\xi}_{k-2}]\le Cb^3$, we have

$$\sup_{x\in\mathsf{R}}\sum_{k=1}^n\mathsf{E}[\widehat{w}_{nk}^2(x)|\widetilde{\xi}_{k-2}]\le Cnb^3.$$

Using the arguments for (5.2) and (5.3), we can show that

$$\sup_{0\le x\le 1}\left|\frac{1}{nb}\sum_{k=1}^n\widehat{w}_{nk}(x)\right|=O_{\mathsf{P}}\left(\sqrt{\frac{b\log n}{n}}\right).$$

The lemma is proved. □

PROOF OF THEOREM 2.5. Write

$$\begin{aligned}\sigma_n^2(x)&=\frac{1}{nhf_{n1}(x)}\sum_{k=1}^n K\left(\frac{X_k-x}{h}\right)[\sigma(X_k)\eta_k]^2\\&\quad+\frac{2}{nhf_{n1}(x)}\sum_{k=1}^n K\left(\frac{X_k-x}{h}\right)[\mu(X_k)-\mu_n(X_k)]\sigma(X_k)\eta_k\\&\quad+\frac{1}{nhf_{n1}(x)}\sum_{k=1}^n K\left(\frac{X_k-x}{h}\right)[\mu(X_k)-\mu_n(X_k)]^2\\&=:\sigma_{n1}^2(x)+c_{n2}(x)+\sigma_{n3}^2(x).\end{aligned}\tag{5.8}$$

We have

$$\begin{aligned}\sup_{0\le x\le 1}|\sigma^2_{n3}(x)| &= O_{\mathsf{P}}\left(\frac{\log n}{nb}+b^4\right)\\ &\quad\times \sup_{0\le x\le 1}\frac{1}{nh}\sum_{k=1}^n\left|K\left(\frac{X_k-x}{h}\right)\right| \\ &= O_{\mathsf{P}}\left(\frac{\log n}{nb}+b^4\right).\end{aligned} \tag{5.9}$$

Using a similar argument as in Zhao and Wu [(2008), page 1875] we have

$$\sup_{0\le x\le 1}|c_{n2}(x)| = O_{\mathsf{P}}\left(\frac{1}{nb^{5/2}}\right). \tag{5.10}$$

For $\sigma^2_{n1}(x)$,

$$\begin{aligned}&(\sigma^2_{n1}(x)-\sigma^2(x))f_{n1}(x)\\ &\quad= \frac{1}{nh}\sum_{k=1}^n K\left(\frac{X_k-x}{h}\right)\sigma^2(x)(\eta_k^2-1)\\ &\qquad+\frac{1}{nh}\sum_{k=1}^n K\left(\frac{X_k-x}{h}\right)(\sigma^2(X_k)-\sigma^2(x))(\eta_k^2-1)\\ &\qquad+\frac{1}{nh}\sum_{k=1}^n K\left(\frac{X_k-x}{h}\right)(\sigma^2(X_k)-\sigma^2(x))\\ &\quad=: M^r_{n2}(x)+R^r_{n2}(x)+R^r_{n3}(x).\end{aligned} \tag{5.11}$$

As in the proof of Lemma 5.3, we get

$$\sup_{0\le x\le 1}|R^r_{n2}(x)| = O_{\mathsf{P}}\left(\sqrt{\frac{b\log n}{n}}\right). \tag{5.12}$$

Also, for $R^r_{n2}(x)$, we have similarly as in Lemma 5.2 that

$$\sup_{0\le x\le 1}|R^r_{n2}(x)-h^2\psi_K\rho_\sigma(x)| = O_{\mathsf{P}}(\tau_n). \tag{5.13}$$

Theorem 2.5 now follows from Lemma 4.4, Proposition 2.1 and (5.8)–(5.13). □

Acknowledgments. We are grateful to two referees and an Associate Editor for their many helpful comments.

REFERENCES

AÏT-SAHALIA, Y. (1996a). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64** 527–560.

AÏT-SAHALIA, Y. (1996b). Testing continuous-time models of the spot interest rate. *Rev. Finan. Stud.* **9** 385–426.

BERMAN, S. (1962). A law of large numbers for the maximum of a stationary Gaussian sequence. *Ann. Math. Statist.* **33** 93–97. MR0133856

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

BLACK, F. and SCHOLES, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81** 637–654.

BOSQ, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction. Lecture Notes in Statistics* **110**. Springer, New York. MR1441072

BRILLINGER, D. R. (1969). An asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.* **75** 545–547. MR0243659

BÜHLMANN, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *Ann. Statist.* **26** 48–83. MR1611804

CHAN, K. C., KAROLYI, A. G., LONGSTAFF, F. A. and SANDERS, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *J. Finance* **47** 1209–1227.

CHAPMAN, D. A. and PEARSON, N. D. (2000). Is the short rate drift actually nonlinear? *J. Finance* **55** 355–388.

COURTADON, G. (1982). The pricing of options on default-free bonds. *J. Finan. Quant. Anal.* **17** 75–100.

COX, J. C., INGERSOLL, J. E. and ROSS, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385–403. MR0785475

CUMMINS, D. J., FILLOON, T. G. and NYCHKA, D. (2001). Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage. *J. Amer. Statist. Assoc.* **96** 233–246. MR1952734

DOUKHAN, P. and LOUHICHI, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–342. MR1719345

DOUKHAN, P., MADRE, H. and ROSENBAUM, M. (2007). Weak dependence for infinite ARCH-type bilinear models. *Statistics* **41** 31–45. MR2303967

DOUKHAN, P. and PORTAL, F. (1987). Principe d'invariance faible pour la fonction de répartition empirique dans un cadre multidimensionnel et mélangeant. *Probab. Math. Statist.* **8** 117–132. MR0928125

DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. MR1997491

ERDÖS, P. (1939). On a family of symmetric Bernoulli convolutions. *Amer. J. Math.* **61** 974–976. MR0000311

FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660. MR1665822

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods.* Springer, New York. MR1964455

FAN, J. and ZHANG, C. (2003). A re-examination of diffusion estimators with applications to financial model validation. *J. Amer. Statist. Assoc.* **98** 118–134. MR1965679

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118. MR0380971

GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1** 15–29. MR0605572

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation From Time Series.* Springer, Berlin. MR1027837

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796. MR1105844

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27** 228–254. MR0971184

Ho, H. C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176. MR0614953

Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Anal.* **12** 402–414. MR0666014

Knafl, G., Sacks, J. and Ylvisaker, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80** 683–691. MR0803261

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883

Neumann, M. H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.* **26** 2014–2048. MR1673288

Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207. MR0732897

Rosenblatt, M. (1976). On the maximal deviation of k-dimensional density estimates. *Ann. Probab.* **4** 1009–1015. MR0428580

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *J. Finance* **52** 1973–2002.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105

Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22** 1328–1345. MR1311978

Tjøstheim, D. (1994). Nonlinear time series: A selective review. *Scand. J. Statist.* **21** 97–130.

Vasicek, O. A. (1977). An equilibrium characterization of the term structure. *J. Financial Economics* **5** 177–188.

Wu, W. B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9** 809–831. MR2047687

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215

Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.* **30** 1441–1459. MR1936325

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 797–811. MR1649488

Zaĭtsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probab. Theory Related Fields* **74** 535–566. MR0876255

Zhao, Z. (2008). Parametric and nonparametric models and methods in financial econometrics. *Stat. Surv.* **2** 1–42. MR2520979

Zhao, Z. and Wu, W. B. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36** 1854–1878. MR2435458

Department of Statistics
University of Pennsylvania
3730 Walnut Street
Philadelphia, Pennsylvania 19104
USA
E-mail: liuweidong99@gmail.com

Department of Statistics
University of Chicago
5734 S. University Avenue
Chicago, Illinois 60637
USA
E-mail: wbwu@galton.uchicago.edu